USGS

Digital Mapping Techniques '03

Association of
American State Geologists

United States
Geological Survey

Digital Mapping Techniques '03— Workshop Proceedings

June 1-4, 2003
Millersville, Pennsylvania

Convened by the
Association of American State Geologists
and the
United States Geological Survey

Hosted by the
Pennsylvania Geological Survey

U.S. GEOLOGICAL SURVEY OPEN-FILE REPORT 03-471

U.S. Department of the Interior
U.S. Geological Survey

Digital Mapping Techniques '03— Workshop Proceedings

Edited by David R. Soller

June 1-4, 2003
Millersville, Pennsylvania

Convened by the
Association of American State Geologists
and the
United States Geological Survey

Hosted by the
Pennsylvania Geological Survey

U.S. GEOLOGICAL SURVEY OPEN-FILE REPORT 03-471
2003

FOREWORD

When one looks at the history of science, the notion of a geologic map is a comparatively new scientific concept. The first geologic map of a large area was produced in England by William Smith, a mere 200 years ago. During a recent visit to the British Geological Survey (BGS), I had the opportunity to learn of their advances in geologic mapping during the last 2 centuries. I was amazed at the current effort to catalog and make all of their mapping holdings digital and more accessible to their various users. The British Isles are perhaps the most geologically mapped part of the planet. However, I was shown the results of an effort to develop a seamless geologic map of Great Britain from their mapped quadrangles. As you would expect, the map looked more like a quilt than a seamless representation of the geology.

The reasons for this problem are many, including varying interpretations of field observations. However, many problems are related to the lack of mapping standards. To me, that fact highlights the importance of this conference and the several that have proceeded it. Clearly, the development of digital mapping standards will serve us well in the future. Consensus and well considered standards will focus our attention on the major scientific issues and differences in interpretation, and will minimize our difficulties in separating legitimate scientific issues from those caused by differences in the standards or methods used to describe geologic features and compile maps. These standards should lead to products that will be far more understandable to the public. William Smith's geologic map was born from very practical needs. I think that even after 200 years, he would be among the first to applaud these efforts and the progress that has been made to date.

The First Digital Mapping Techniques Workshop was held in 1997 in Lawrence, Kansas, with the goal of "developing more cost-effective, flexible, and useful systems for digital mapping and GIS analysis". This joint effort of the Association of American State Geologists (AASG) and the U.S. Geological Survey (USGS) has been a great success, and instrumental in the sharing of ideas, technologies, and methods that will truly contribute to the development of digital geologic map products and standards designed to better serve society's needs in the future. It was a great pleasure to welcome the participants of the sixth annual workshop in Millersville, Pennsylvania. Enormous progress has been made since the first workshop, but much work remains. This volume, like those from previous workshops that document the 'state of the science' in digital mapping techniques, is an important legacy of the outstanding progress being made.

P. Patrick Leahy
U.S. Geological Survey
Associate Director for Geology

CONTENTS

Foreword
By Pat Leahy (U.S. Geological Survey) III

Introduction
By David R. Soller (U.S. Geological Survey). 1

Oral Presentations

Bedrock Geology and Bedrock Topography GIS of Ohio: New Data and Applications for Public Access
By James McDonald, E. Mac Swinford, Lawrence H. Wickstrom, Ernie R. Slucher, Donovan M. Powers, and Thomas M. Berg (Ohio Geological Survey) . . 3

Digital Karst Density Layer and Compilation of Mapped Karst Features in Pennsylvania
By Stuart O. Reese and William E. Kochanov (Pennsylvania Geological Survey) . . 17

Southern California Areal Mapping Project (SCAMP) and Multidimensional Databases
By Douglas M. Morton (U.S. Geological Survey) 23

Digital Geologic Mapping in a Data Rich, Urban Environment
By Gayle H. "Scott" McColloch and Jane S. McColloch (West Virginia Geological and Economic Survey) 27

Alaska Digital Mapping Project: From Field to Publication
By Christopher P. Garrity (U.S. Geological Survey). 33

A Map of Lightning Strike Density for Southeastern Pennsylvania, and Correlation with Terrain Elevation
By Alex J. DeCaria and Michael J. Babij (Millersville University) 37

Negligence and Professional Malpractice Related to GIS Datasets
By Ian J. Duncan (Virginia Division of Mineral Resources). 41

From CARIS GEMM 4 to *Carta for Geology* Software: Our Responses to Geology Educational Challenges in the Age of Digital Mapping
By H. Wouter van de Poll (Univ. of New Brunswick) and Chris Parsons (CARIS) . . 47

The National Geologic Map Database: Overview and Progress
By David R. Soller (U.S. Geological Survey) and Thomas M. Berg (Ohio Geological Survey) 57

CHRONOS—Integrated Chronostratigraphic Databases, the Development of an International Standard Time Scale and the Interoperability with Time Scales of U.S. State Surveys
By Bruce R. Wardlaw (U.S. Geological Survey) 79

The National Geologic Map Database Image Library—General Concepts
By David R. Soller (U.S. Geological Survey) and Thomas M. Berg (Ohio Geological Survey) 89

The National Geologic Map Database Image Library—Technical Details
By Robert S. Wardwell, Kevin W. Laurent, Jeremy O. Skog, and David R. Soller (U.S. Geological Survey) 95

Science Language for Geologic-map Databases in North America: A Progress Report
By NADM Science Language Technical Team. 109

Conversion of Lithological Data in the Manitoba Water Well Database (GWDrill) to a Mappable Format
By L. H. Thorleifson and M. Pyne (Geological Survey of Canada) 139

Geoscience Terminology Development for the National Geologic Map Database
By Steven M. Richard (Arizona Geological Survey/U.S. Geological Survey), Jonathan C. Matti, and David R. Soller (U.S. Geological Survey) 157

Geologic Map Database Implementation in the ESRI Geodatabase Environment
By Stephen M. Richard (Arizona Geological Survey/U.S. Geological Survey). . 169

Portable Software Tools for Managing and Referencing Taxonomies
By Jordan Hastings (U.S. Geological Survey and U.C., Santa Barbara) 185

CGMW Working Group on Standards for Digital Geological Data ("DIMAS")
By Kristine Asch (BGR—German Federal Institute for Geosciences and Natural Resources) 193

Preservation of Geoscience Data and Collections
By Linda R. Musser (Pennsylvania State University). 195

Integrating the Data Repository with the Publication Process
By Andrew Moore (Geological Survey of Canada), Linda Guay (ESSInfo, Natural Resources Canada), and Ross Murray (Geological Survey of Canada). . . 197

The Alaska Division of Geological & Geophysical Survey's Metadata Policy Development and Implementation
By Carrie L. Browne and Larry K. Freeman (Alaska Division of Geological & Geophysical Survey), and Gina R.C. Graham 201

National Spatial Data Clearinghouse: The Real Story
By Peter N. Schweitzer (U.S. Geological Survey) 209

XML Encoding of the North American Data Model
By NADM Data Interchange Technical Team. 215

GEON: Toward a Cyberinfrastructure for the Geosciences—A Prototype for Geologic Map Integration via Domain Ontologies
By Bertram Ludäscher and Kai Lin (San Diego Supercomputer Center), Boyan Brodaric (Geological Survey of Canada), and Chaitan Baru (San Diego Supercomputer Center) 223

Poster Presentations

Delaware Inland Bays Shoreline Extraction Using Landsat 7 Satellite Imagery
By Lillian T. Wang (Delaware Geological Survey). 231

Applying Surficial Geologic Mapping to Geologic Hazards Mapping, Nez Perce County, Idaho
By Loudon R. Stanford and Kurt L. Othberg (Idaho Geological Survey), Bill Reynolds (Nez Perce County), and Roy M. Breckenridge (Idaho Geological Survey) 235

Creating a Shaded-Relief Geologic Map Using World Construction Set and Adobe Illustrator Software
By Loudon R. Stanford (Idaho Geological Survey) 237

Selection of an Appropriate Base Map for a Statewide Geologic Mapping Program
By Jane E. Domier and Donald E. Luman (Illinois State Geological Survey) . . . 239

Kentucky Land-Use Planning on the Web
By Gerald A. Weisenfluh and Daniel I. Carey (Kentucky Geological Survey). . . 245

The National Park Service Geologic Resources Inventory "From Paper to Digital: Exploring a Geologic-GIS Map"
By Stephanie A. O'Meara, Trista L. Thornberry (Colorado State University/ National Park Service Cooperator), Anne R. Poole (National Park Service), and Victor deWolfe (Colorado State University/National Park Service Cooperator) 247

Meeting Constituent Needs for 1:24,000 Scale Geoscience Data in New Jersey
By Ronald Pristas (New Jersey Geological Survey) 255

Problems and Solutions in the Digital Compilation and Production of the "Map of Surficial Deposits and Materials in the Eastern and Central United States (East of 102° West Longitude)"
By Charles A. Bush, Diane E. Lane, David S. Fullerton, and Nancy Shock (U.S. Geological Survey) 257

Appendix. List of Workshop Attendees 261

Introduction

By David R. Soller

U.S. Geological Survey
926-A National Center
Reston, VA 20192
Telephone: (703) 648-6907
Fax: (703) 648-6977
e-mail: drsoller@usgs.gov

The Digital Mapping Techniques '03 (DMT'03) workshop was attended by nearly 90 technical experts from 36 agencies, universities, and private companies, including representatives from 22 state geological surveys (see Appendix A). Although the meeting was slightly smaller than DMT'02 it was, considering the budget deficits in nearly all 50 states, very well attended. This workshop was similar in nature to the previous six meetings, held in Lawrence, Kansas (Soller, 1997), in Champaign, Illinois (Soller, 1998), in Madison, Wisconsin (Soller, 1999), in Lexington, Kentucky (Soller, 2000), in Tuscaloosa, Alabama (Soller, 2001), and in Salt Lake City, Utah (Soller, 2002). This year's meeting was hosted by the Pennsylvania Geological Survey, from June 1-4, 2003, on the Millersville University campus in Millersville, Pennsylvania. As in the previous meetings, the objective was to foster informal discussion and exchange of technical information. This objective was well met, as attendees continued to share and exchange knowledge and information, and to renew friendships and collegial work begun at past DMT workshops.

All the DMT workshops have been coordinated by the Association of American State Geologists (AASG) and U.S. Geological Survey (USGS) Data Capture Working Group, which was formed in August 1996, to support the AASG and the USGS in their effort to build a National Geologic Map Database (see Soller and Berg, this volume, and <http://ncgmp.usgs.gov/ngmdbproject/standards/datacapt/>). The Working Group was formed because increased production efficiencies, standardization, and quality of digital map products were needed for the database—and the State and Federal geological surveys—to provide more high-quality digital maps to the public.

At the 2003 meeting, oral and poster presentations and special discussion sessions emphasized 1) methods for creating and publishing map products (here, "publishing" includes Web-based release); 2) digital cartographic techniques, 3) analytical GIS techniques; 4) continued development of the National Geologic Map Database; 5) progress toward building and implementing a standard geologic map data model and standard science language, and 6) the need to archive both the published products and the data and observational data that support it.

ACKNOWLEDGMENTS

I thank the Pennsylvania Geological Survey (PGS) and their Director and State Geologist, Jay Parrish, for hosting this meeting and for his enthusiastic participation. In the tradition of past DMT meetings, the attendees were given a very informative, productive, and enjoyable experience. I especially thank Tom Whitfield (PGS), who coordinated the events, provided excellent support for the attendees, and offered an entertaining range of technical and social activities, including a medieval-themed dinner. Thanks also to Lynn M. Goodling (PGS) for skillfully managing the registration and logistics, Stuart O. Reese (PGS) for managing the meeting's website, and Gary M. Fleeger (PGS and secretary of the Field Conference of Pennsylvania Geologists) for hosting the meeting's Web site and finances. I also warmly thank Millersville University for providing an excellent venue and support for our meeting; in particular, Dr. Richard D. Clark, Dr. Lynn Marquez, and Dr. Alex DeCaria (Department of Earth Sciences), and John M. Roscoe (Director of University Dining and Conference Services).

I also, with gratitude, acknowledge Tom Berg (Chair, AASG Digital Geologic Mapping Committee) for his friendship and his help in conducting the meeting, and for his continued support of AASG/USGS efforts to collaborate on the National Geologic Map Database. Thanks of course also are extended to the members of the Data Capture Working Group (Warren Anderson, Kentucky Geological Survey; Rick Berquist and Elizabeth Campbell, Virginia Division of Mines and Geology; Rob Krumm and Barb Stiff, Illinois State Geological Survey; Scott McColloch, West Virginia Geological and Economic Survey; Gina Ross, Kansas Geological Survey; George Saucedo, California Geological Survey; and Tom Whitfield, Pennsylvania Geological Survey) for advice in planning the workshop's content.

I warmly thank Kit Richter (Geological Survey of Alabama) for technical editing of each manuscript, and Lisa Van Doren (Ohio Geological Survey) for typesetting the Proceedings. Finally, I thank all attendees for their participation; their enthusiasm and expertise were the primary reasons for the meeting's success.

PRESENTATIONS

The workshop included 29 oral presentations. Nearly all are supported by a short paper contained in these Proceedings. The papers describe technical and procedural approaches that currently meet some or all needs for digital mapping at the respective agency. There is not, of course, a single "solution" or approach to digital mapping that will work for each agency or for each program or group within an agency; personnel and funding levels, and the schedule, data format, and manner in which we must deliver our information to the public require that each agency design their own approach. However, the value of this workshop and other forums like it is through their roles in helping to design or refine these agency-specific approaches to digital mapping, and to find applicable approaches used by other agencies. In other words, communication helps us to avoid "reinventing the wheel."

POSTERS AND COMPUTER DEMOS

More than 20 posters were exhibited and several computer demonstrations were provided throughout the workshop. These provided an excellent focus for technical discussions and support for oral presentations. Many are documented with a paper in these Proceedings, following those for the oral presentations; the other posters generally provided material in support of oral presentations, and so are not documented here.

DISCUSSION SESSIONS

To provide the opportunity to consider a topic in some detail, special discussion sessions are held at the DMT workshops. This year there were two: 1) how we can share information about digital cartographic techniques, and 2) how we should try to archive the basic data and observations that form the basis for all published databases, maps and summaries. Discussion on these topics was concentrated on the final day of the meeting, and produced many good ideas and recommendations that will be discussed by the Data Capture Working Group and DMT'03 attendees via the DMTListserve. These sessions highlight an important aspect of the DMT workshop series—it provides a unique venue for sharing technical information and experience for those in the geologic and GIS disciplines.

THE NEXT DMT WORKSHOP

The eighth annual DMT meeting will be held in mid-May, 2004, in Portland, Oregon. Please consult the Web site <http://ncgmp.usgs.gov/ngmdbproject/standards/datacapt/> for updated information. While planning for that event, the Data Capture Working Group will carefully consider the recommendations offered by DMT'03 attendees.

REFERENCES

Soller, D.R., ed., 1997, Proceedings of a workshop on digital mapping techniques: Methods for geologic map data capture, management, and publication: U.S. Geological Survey Open-File Report 97-269, 120 p., <http://ncgmp.usgs.gov/pubs/of97-269/>.

Soller, D.R., ed., 1998, Digital Mapping Techniques '98—Workshop Proceedings: U.S. Geological Survey Open-File Report 98-487, 134 p., <http://pubs.usgs.gov/openfile/of98-487/>.

Soller, D.R., ed., 1999, Digital Mapping Techniques '99—Workshop Proceedings: U.S. Geological Survey Open-File Report 99-386, 216 p., <http://pubs.usgs.gov/openfile/of99-386/>.

Soller, D.R., ed., 2000, Digital Mapping Techniques '00—Workshop Proceedings: U.S. Geological Survey Open-File Report 00-325, 209 p., <http://pubs.usgs.gov/openfile/of00-325/>.

Soller, D.R., ed., 2001, Digital Mapping Techniques '01—Workshop Proceedings: U.S. Geological Survey Open-File Report 01-223, 248 p., <http://pubs.usgs.gov/of/2001/of01-223/>.

Soller, D.R., ed., 2002, Digital Mapping Techniques '02—Workshop Proceedings: U.S. Geological Survey Open-File Report 02-370, 214 p., <http://pubs.usgs.gov/of/2002/of02-370/>.

Bedrock Geology and Bedrock Topography GIS of Ohio: New Data and Applications for Public Access

By James McDonald, E. Mac Swinford, Lawrence H. Wickstrom, Ernie R. Slucher, Donovan M. Powers, and Thomas M. Berg

Ohio Division of Geological Survey
4383 Fountain Square Drive
Columbus, OH 43224
Telephone: (614) 265-6601
Fax: (614) 447-1918
e-mail: {jim.mcdonald, mac.swinford, larry.wickstrom, ernie.slucher, donovan.powers, thomas.berg}@dnr.state.oh.us

INTRODUCTION

The Ohio Division of Geological Survey (ODGS) recently completed four major geographic information systems (GIS) data sets: bedrock geology and bedrock topography at 1:24,000 scale, and bedrock geology and bedrock topography at 1:500,000 scale. These data sets, the result of several geologic mapping cooperative programs between ODGS and various state and federal agencies, represent a major advance for the ODGS and its goal of providing digital geologic information for the general public.

This summary briefly describes the history and procedures used to map the bedrock geology and bedrock topography of Ohio and the conversion of the open-file 1:24,000-scale bedrock geology and bedrock topography maps into GIS data layers. The 1:24,000-scale bedrock geology and bedrock topography data sets form the framework for the new 1:500,000-scale bedrock geology and bedrock topography maps and GIS data layers. In addition, computer applications were developed to plot the maps and export GIS data layers for individual 7.5-minute quadrangles. These applications will allow the ODGS to manage and update the GIS data sets and will assist in the distribution of the maps and GIS data sets to the public.

GEOLOGIC MAPPING IN OHIO

Prior to the completion of the new bedrock geology maps and GIS data sets, Ohio's 1920-vintage state bedrock geology map was one of the oldest in-print state bedrock geology maps in the nation (Bownocker, 1920). Mapping of Ohio's bedrock geology has been going on since the release of the 1920 map, but had not resulted in the production of a new statewide bedrock geology map. From 1918 to 1979, the ODGS conducted geologic mapping on the county level at scales of 1:62,500 and smaller, on base maps constructed from the U.S. Geological Survey (USGS) 15-minute topographic quadrangles (Sherman, 1933). Only 19 of the 88 counties had been completed by the time the last county report (Cuyahoga County) was published (Ford, 1987). Starting in the late 1960s, a more detailed level of mapping was initiated. Between 1957 and 1963, a new topographic map series was created for Ohio at a scale of 1:24,000 by the USGS (Bernhagen, 1994). With the completion of that program, there were a number of initiatives, starting in the mid-1960s (e.g. DeLong, 1965), to perform detailed, field-based geologic mapping at 1:24,000 scale. By the end of the 1980s, only 37 quadrangles, out of 788, had been completed and it became apparent that, with the current staffing level, completing the new detailed bedrock geology mapping effort for the entire state would take more than 100 years.

With the appointment of a new state geologist in 1989, a program was initiated at the ODGS to perform more rapid reconnaissance-level geologic mapping at 1:24,000 scale, funded by a portion of the Ohio Minerals Severance Tax and supplemented by funds from USGS COGEOMAP and STATEMAP grants, U.S. Environmental Protection Agency (EPA) Nonpoint-Source Pollution Program 319(h) grants, and grants from the Ohio Department of Transportation (ODOT). The reconnaissance approach to bedrock geology mapping relies less on detailed, field-based procedures and more on in-house data and analysis. The reconnaissance-mode techniques included limiting the stratigraphic resolution of the mapped units, the use of data that was collected over the last 120 years and stored in the ODGS archives, the use of existing bedrock topography maps, and the use of computers to assist in the mapping. These techniques, while admittedly reducing the possible detail, enabled the entire state to be mapped in 7 years, a fairly remarkable timeframe for such a large project (Swinford, 1997; McDonald, 2002).

Several methods were used to limit the stratigraphic resolution of the reconnaissance-level bedrock geology maps. For some stratigraphic intervals, mapping units were combined. For example, the Silurian Peebles Dolomite, Lilly Formation, and Bisher Formation were too thin to be mapped as individual units at 1:500,000 scale. These formations were combined to form a single unit—Peebles Dolomite, Lilly Formation, and Bisher Formation undivided. Some stratigraphic intervals, such as the Ordovician undifferentiated and the Mississippian undifferentiated, were not formally defined. In the case of the Ordovician undifferentiated, the physical characteristics and spatial variability of this lithologic unit, buried beneath glacial deposits, were too poorly understood to be formally defined. In another example, the Pottsville, Allegheny, Conemaugh, and Monongahela Groups of the Pennsylvanian System and the Dunkard Group of the Pennsylvanian-Permian Systems were not subdivided below the group level, in part because few of the internal units have any lateral continuity (Larsen, 1991). As a final example, individual coal beds were not mapped, except where they were used as key beds to define group boundaries. While each of these methods reduced the stratigraphic resolution of the mapping units, the process also decreased the amount of time needed to resolve stratigraphic problems and helped expedite the mapping.

There was a limited amount of fieldwork performed for this project. The data, some of which dated back 120 years, primarily came from the extensive ODGS archives. The data collection in the ODGS archives contains in excess of 17,000 measured stratigraphic sections, 4,000 core descriptions, 250,000 oil- and gas-well records, and many geologists' field notebooks. Most of the measured stratigraphic section and core descriptions have been entered into the USGS's National Coal Resources Data System (NCRDS). These NCRDS data points, together with the oil- and gas-well geophysical logs, were used to build cross sections and to identify the tops of mappable units in the subsurface. Where necessary, field work was performed to collect additional data to fill gaps and to resolve complex stratigraphic and structural issues. Eleven stratigraphic test core holes were drilled in western Ohio by ODGS to supplement areas of sparse data. These core holes were geophysically logged to allow correlation to nearby oil and gas wells.

Because nearly three-quarters of Ohio is covered by glacial and related deposits, the bedrock topography had to be mapped before the bedrock geology. The bedrock topography data set was created using many different sources. There has been an ongoing effort to map the bedrock topography since the early 1970s. At that time, the ODGS started mapping the bedrock topography on a county-by-county basis at 1:62,500 scale. In the mid-to-late 1980s, the ODGS began more detailed bedrock topography mapping on 7.5-minute quadrangles at 1:24,000 scale. In both instances, the bedrock topography was mapped using available water-well logs from the Ohio Division of Water, ODOT boring logs, and other data sources. For the reconnaissance bedrock geology mapping, the existing 1:62,500-scale county maps were photographically enlarged to 1:24,000 scale. In areas where there were no bedrock topography maps, new maps were created at 1:24,000-scale (figure 1).

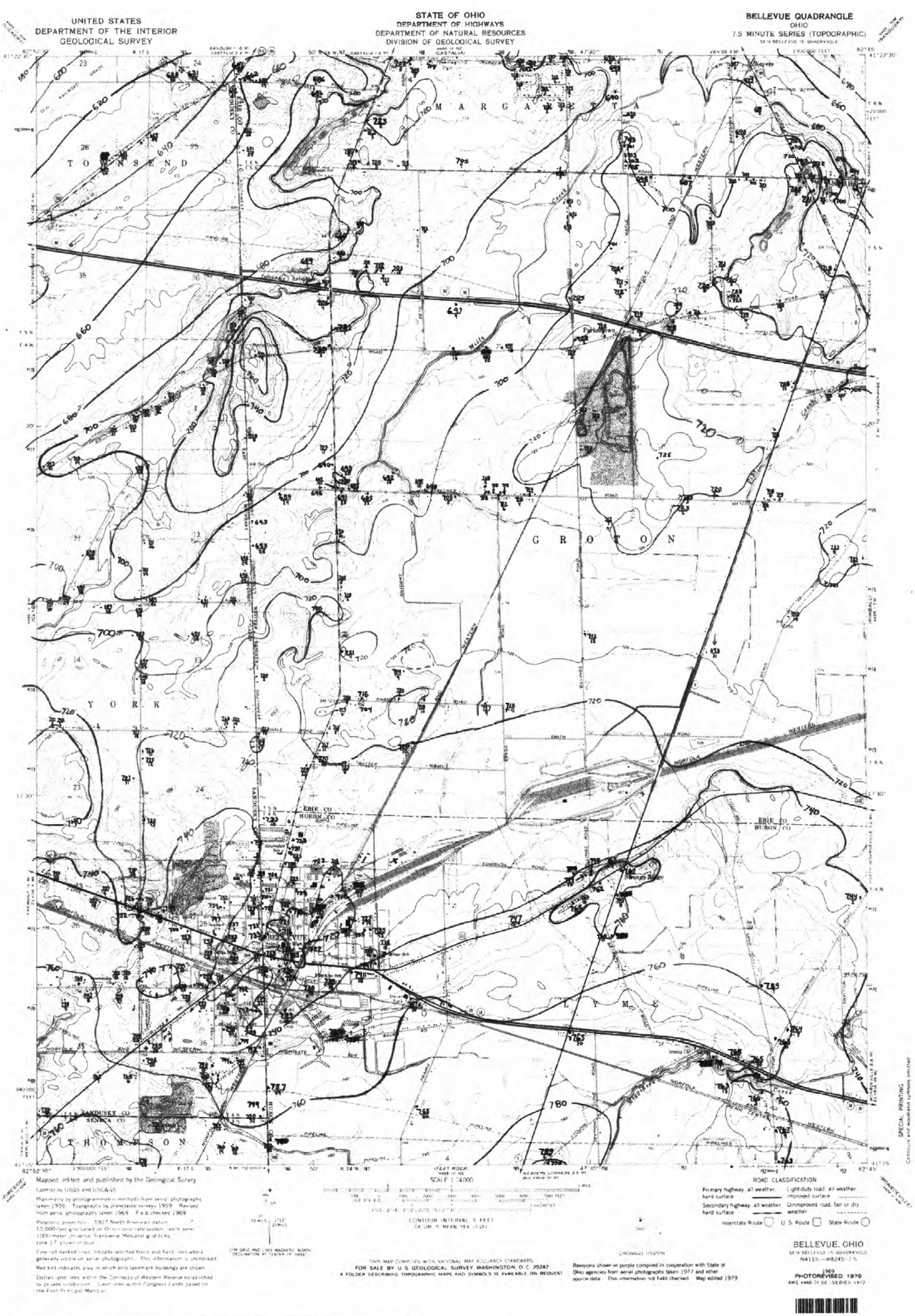
The ODGS had just begun to use computers in the mid-to-late 1980s. In order to accelerate the pace of mapping, the ODGS in 1991 acquired PC-based gridding and contouring software to create structure contour maps of the mappable units in each quadrangle. Depending on how many contacts were present in the area at the surface or beneath the glacial drift, between one and eight different structure contour maps were drawn per 7.5-minute quadrangle (figure 2) (McDonald, 2002).

Structure contour maps were used in conjunction with bedrock topography maps, drawn on USGS 7.5-minute topographic maps, to produce 1:24,000-scale bedrock geology maps. The structure contour maps were overlaid on hand-drawn bedrock topography maps which, in turn, were overlain by mylar 1:24,000-scale base maps. Bedrock contacts were drawn on the base maps where the structure contour surface intersected equivalent elevations on the bedrock topography or the surface topography maps, producing new bedrock geology maps (figure 3). By the end of the project, the geologists had created over 1,840 structure contour maps, 788 bedrock topography maps, and 788 bedrock geology maps (McDonald, 2002).

1:24,000-SCALE BEDROCK GEOLOGY AND BEDROCK TOPOGRAPHY GIS

After the bedrock geology maps were drawn, the maps were converted to a GIS. Funding for GIS conversion came from the State of Ohio General Revenue Fund, supplemented by grants from ODOT and a USGS STATEMAP grant. The conversion was also abetted by two other factors. During the mid-1990s, the Ohio Office of Budget and Management allowed the use of capital funds, which are normally used for building and maintaining roads and buildings, for building digital data infrastructure elements. These elements included the conversion of data to a GIS, using outside contractors to perform the conversion.

The second factor was the initiation of a CAD/GIS conversion shop by the Ohio Penal Industries (Wickstrom, McDonald, and Berg, 1998). Using inmates trained in CAD and GIS software as a labor force, and private-sector contractors as project management, Ohio Penal Industries provided map digitizing services at a reasonable cost. ODGS was pleased by the quality of work provided by the inmates, who were motivated and took pride in the products they created. The work afforded them an



STATE OF OHIO, DEPARTMENT OF NATURAL RESOURCES, DIVISION OF GEOLOGICAL SURVEY

OPEN-FILE MAP: BT-D3C7
TITLE: BEDROCK TOPOGRAPHY OF THE BELLEVUE, OHIO, QUADRANGLE.
Contour Interval: 20 Feet
•745 Water Well Location and Elevation of Bedrock Surface.
Date completed: AUGUST 1995 Geology by: Larsen 1984 1994; Pavey, 1986
Date Fieldwork conducted: ________ to ________ DRAFTED BY J. KRIZ, 1995

The information portrayed/contained herein may be preliminary and subject to revision. The Ohio Geological Survey has technically reviewed this document but disclaims any responsibility or liability for interpretations or decisions based thereon.

Prepared in cooperation with the Department of the Interior
United States Geological Survey
National Geologic Mapping Act of 1992 -
STATEMAP Component

— Revisions made by G. Larsen
8/1999

Figure 1. A bedrock topography contour map of the Bellevue, Ohio 7.5-minute quadrangle, as released in the ODGS informal, open-file series.

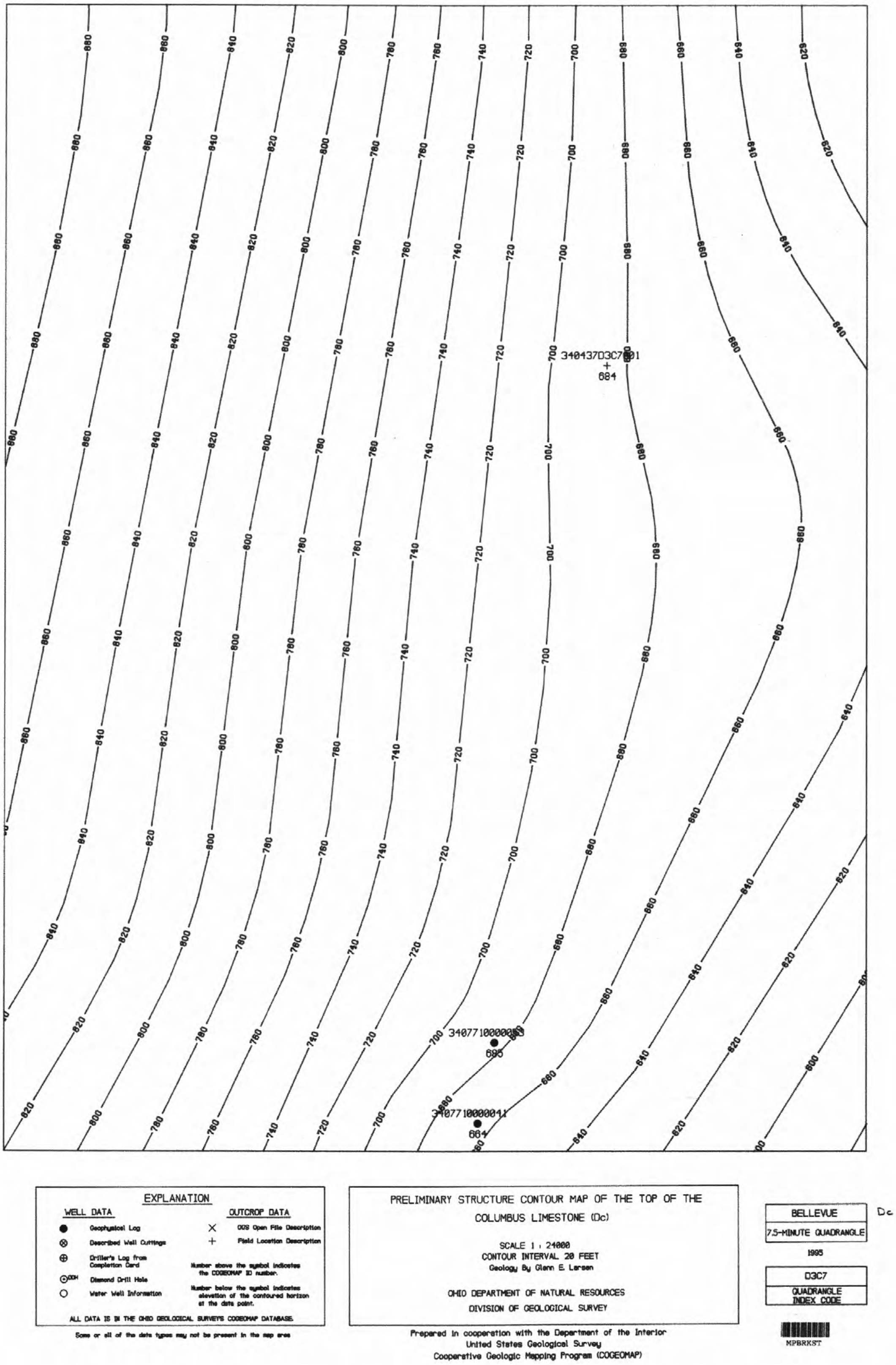


Figure 2. A computer-generated structure contour map, as released in the ODGS informal, open-file series. This is a structure contour map on the top of the Devonian Columbus Limestone of the Bellevue, Ohio 7.5-minute quadrangle.

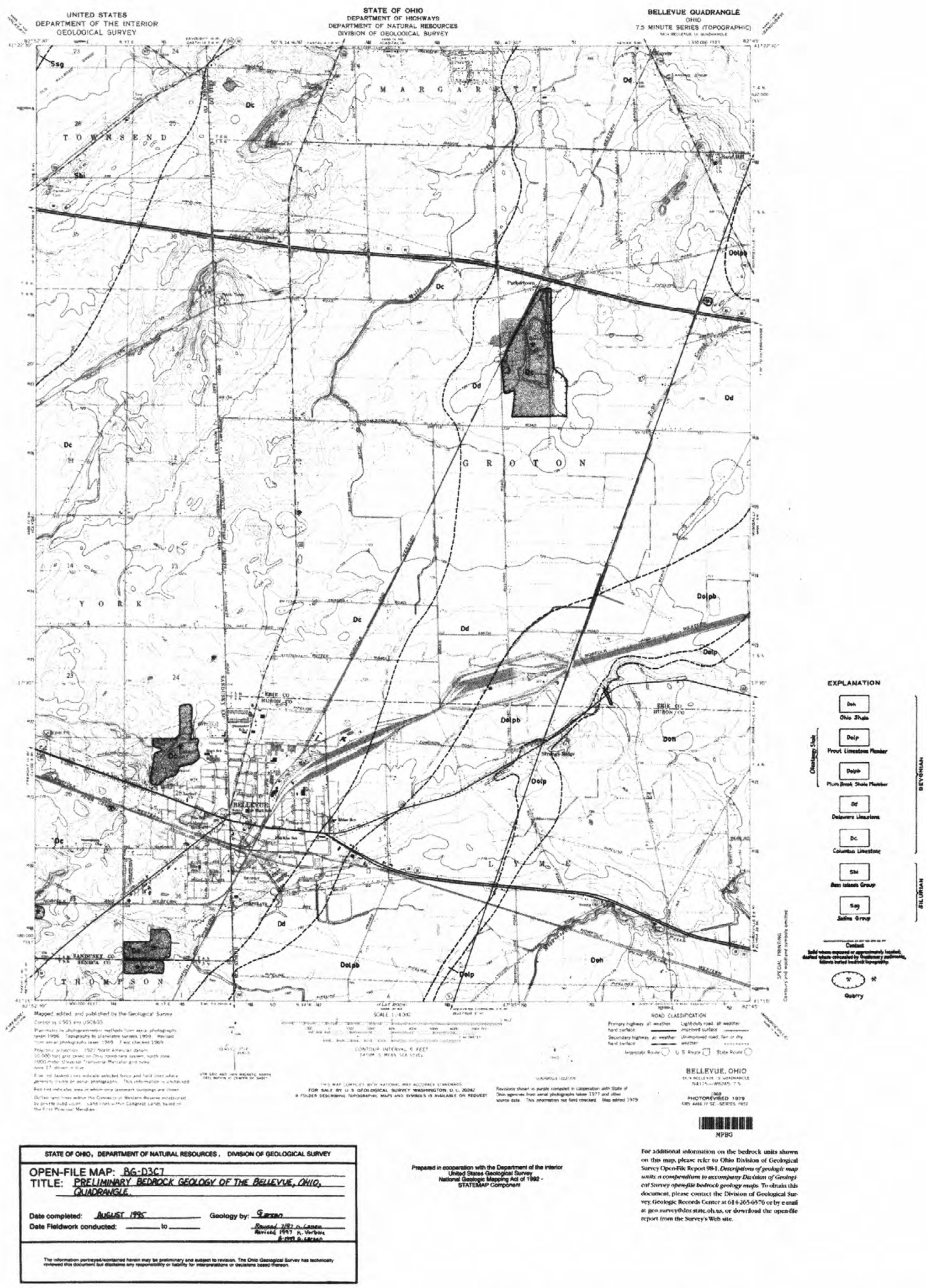


Figure 3. A bedrock geology map of the Bellevue, Ohio 7.5-minute quadrangle, as released in the ODGS informal, open-file series.

opportunity to acquire meaningful training in the use of computers and software. Non-tangible incentives for the prison inmates assisted in keeping job satisfaction and quality of the GIS product at a high level.

The conversion of the bedrock geology and bedrock topography maps was performed over a 10-year period. Initially, ODGS staff digitized the bedrock geology maps of western and part of southern Ohio between 1992 and 1998. The ODGS staff and interns digitized the bedrock geology maps of the central part of the state under a USGS grant in 1996. Finally, the Ohio Penal Industries digitized the bedrock geology maps of the eastern part of the state in 1998 (figure 4). The ODGS staff digitized the bedrock topography maps for northwestern Ohio in 1992 and the central part of the state in 1996. The Ohio Penal Industries digitized the remaining bedrock topography maps, in southwestern and eastern Ohio, in 2001-2002.

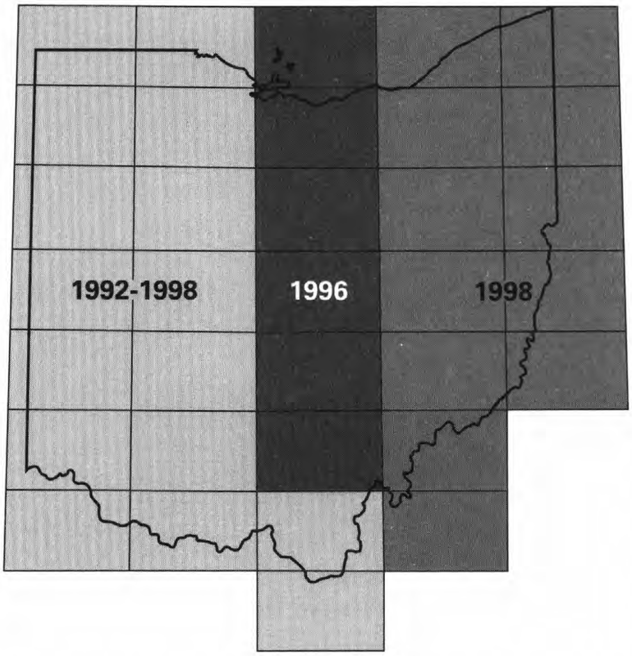


Figure 4. Index map showing the years during which the bedrock geology maps were digitized.

Digitization of bedrock geology and bedrock topography data sets was accomplished using digitizing tables and heads-up digitizing techniques. ODGS staff and interns digitized the 7.5-minute quadrangle maps on a digitizing table, registering the map to a digital file containing the 7.5-minute quadrangle boundaries. The staff and interns then digitized the contact lines into the bedrock geology and bedrock topography CAD files. The inmates performed their work differently, first scanning the mylar maps and then registering the images to the correct quadrangle coordinates. They then did heads-up digitizing of the 7.5-minute quadrangle bedrock geology and bedrock topography maps.

The private-sector contractors provided project management and software programming, as well as training, guidance, and quality control for the inmates. The final project-management contractor also converted the CAD files to seamless ArcInfo coverages. The GIS files were then sent to ODGS for final quality control and acceptance, and were inspected by an ODGS college intern to ensure that the graphics and the polygon and line attributes were correctly captured. The GIS conversion was completed in 2002.

GENERALIZED BEDROCK GEOLOGY AND BEDROCK TOPOGRAPHY GIS

The new full color 1:500,000-scale state bedrock geology map (figure 5) was produced from geologic contacts compiled, generalized, and originally drawn at 1:250,000 scale. First, digital CAD files of edge-matched 7.5-minute (1:24,000) bedrock geology quadrangle maps were plotted onto mylar as 1:250,000-scale, 1 x 2-degree quadrangles. These were then overlaid by a second sheet of mylar containing reverse plots of the digital state transportation network (provided by ODOT) and a modified version of the digital drainage network (USGS). This second sheet of mylar served as the base map for all compilations.

Printing these base map elements as reverse features (upside down and backwards) on the back side or bottom of two-sided mylar served two functions. First, although these plots would be on the back side of the mylar, they would appear geographically correct when placed right-side-up on a light table. Second, the top side of this same sheet of mylar could then serve as the drawing surface for the geologic contact phase of map compilation. This method was used so that any drawing errors made during the compilation phase could be corrected without destroying the transportation and drainage network features plotted on the back side of the base map.

Finally, all contact lines between mapped geologic units were duplicated and some were modified locally by hand. Generalizing and modification of these lines also assured that the hand drawn contact lines maintained their proper spatial relationships to the hydrology and transportation network elements on the map.

The generalized 1:250,000-scale contact lines were then scanned, registered to the new digital state transportation map, heads-up digitized into a CAD file, and converted to an ArcInfo coverage. Topology was created for these elements, with a fuzzy tolerance appropriate for a 1:250,000-scale map. Labels within the coverage area were attributed with the correct geologic unit abbreviation; individual (1 x 2-degree sheets) coverage areas were merged; and to finish the process, the boundaries between the individual map sheets were dissolved.

The generalized bedrock geology data set is com-

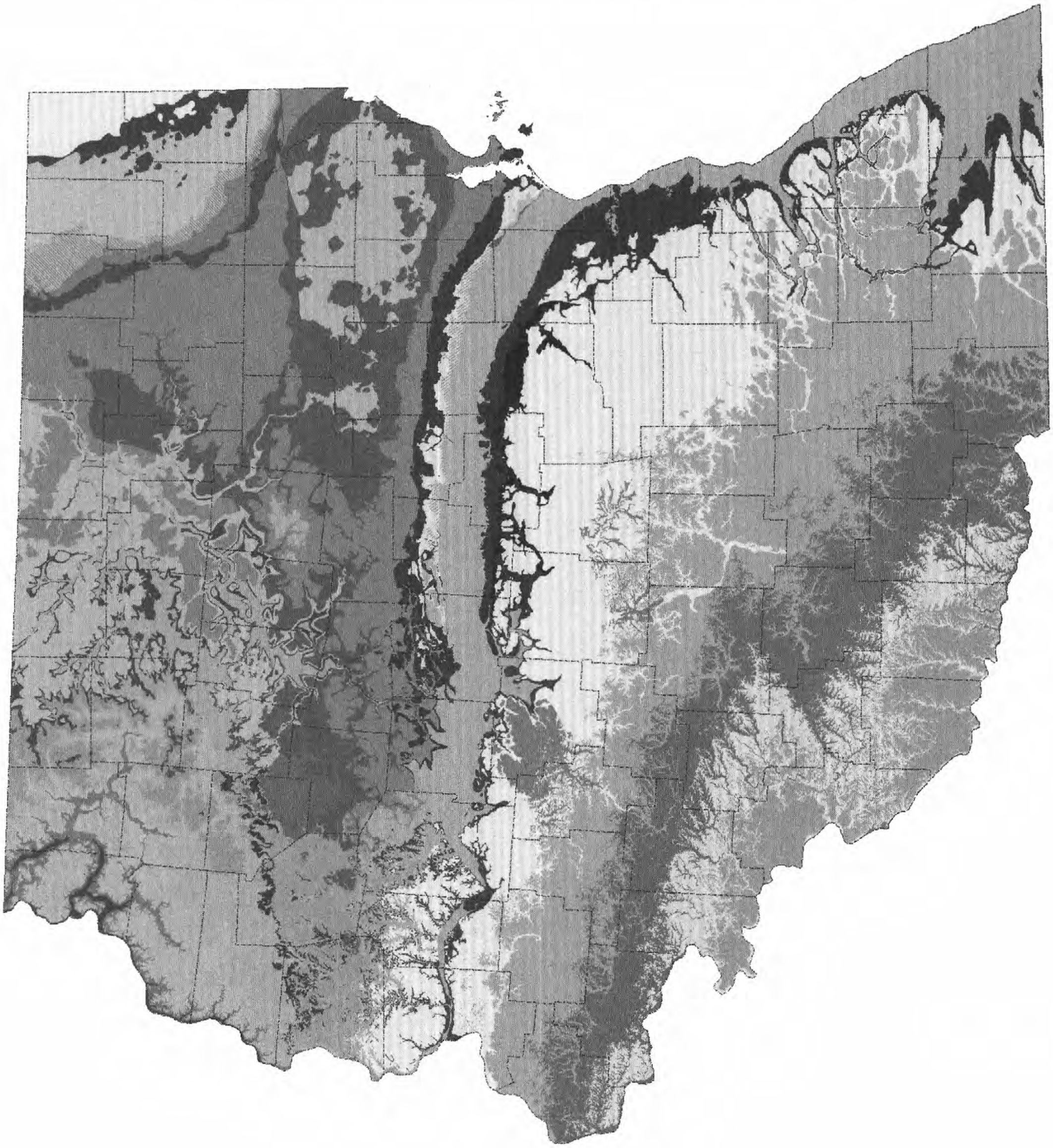


Figure 5. A view of the new state bedrock geology map, to be published at 1:500,000 scale.

posed of four separate GIS data layers. The geologic-unit polygons and contact lines are the two largest GIS data layers. The other two data layers contain structural elements (faults) and pinch-out lines. The term pinch-out lines is a special connotation developed during the project to designate those areas on the map where, because of high topographic relief or lithologic thinning of a geologic unit, two adjacent geologic contact lines between different map units were merged together to form a single line. On the final 1:500,000-scale bedrock geology map of Ohio, these lines will be depicted by a heavy line that is three to four times the width of a normal geologic-contact line.

ODGS has employed some unique features in compiling the 1:500,000-scale bedrock geology GIS for the creation of the final publication layout. The cross section and correlation chart that accompany the actual map are also GIS data layers. Typically, layouts to publish geologic maps use raster images of cross sections, correlation charts, and stratigraphic columns. Instead, the cross section (figure 6) and correlation chart (figure 7) were digitized in relative units and converted to GIS data layers. These two data layers were inserted into the map layout and linked to the attribute table so they can be symbolized the same way as, and in conjunction with, the 1:500,000-scale map layers.

A 1:500,000-scale bedrock topography GIS data layer and plot-on-demand map also were compiled digitally by the ODGS. Compilation of the statewide bedrock topography layer presented unique problems because of the different vintages of the original maps and the variety of contour intervals (5, 10, 20, 25, 50, and 100 feet) that had to be used across the state depending on the gradient of

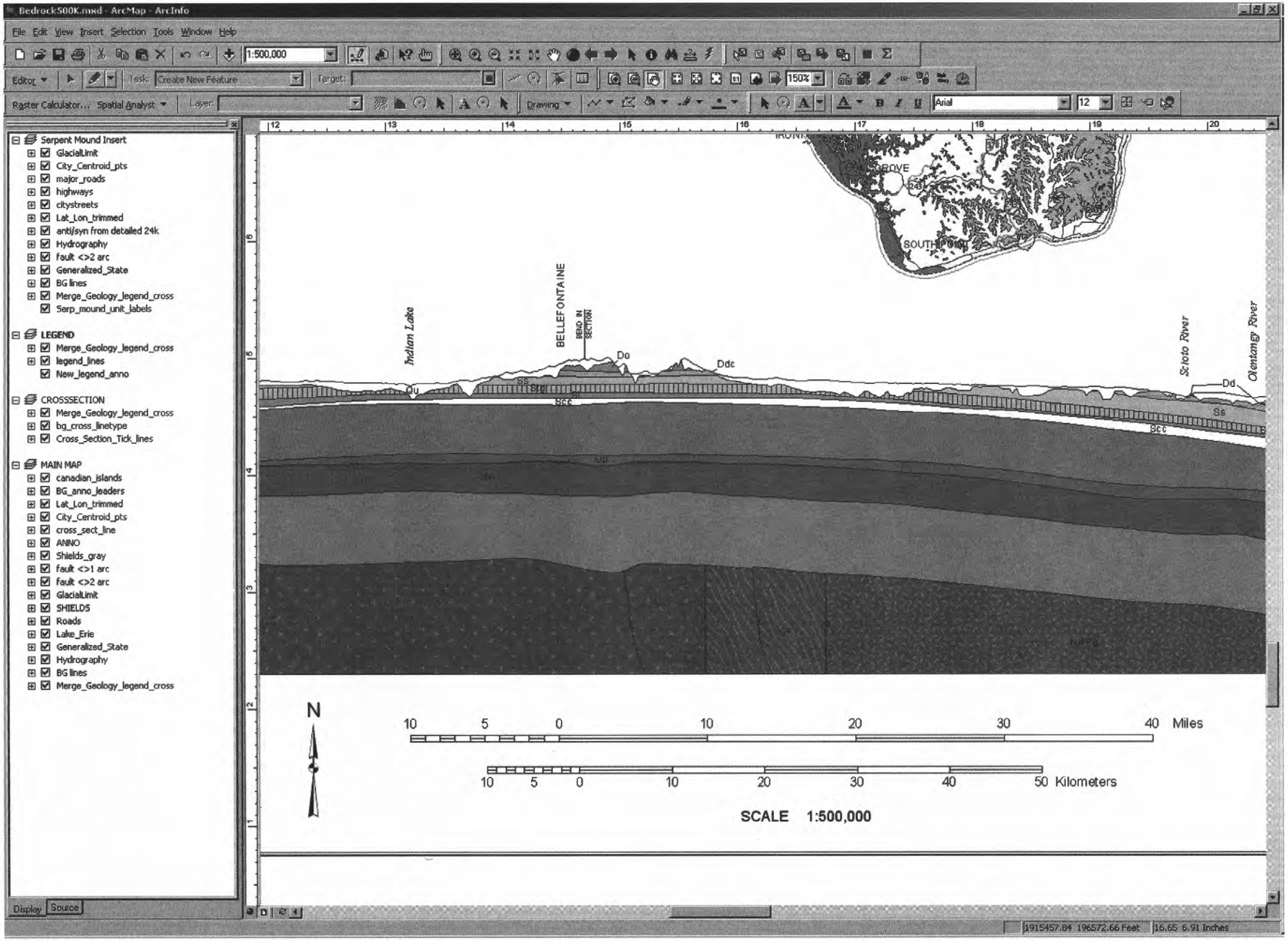
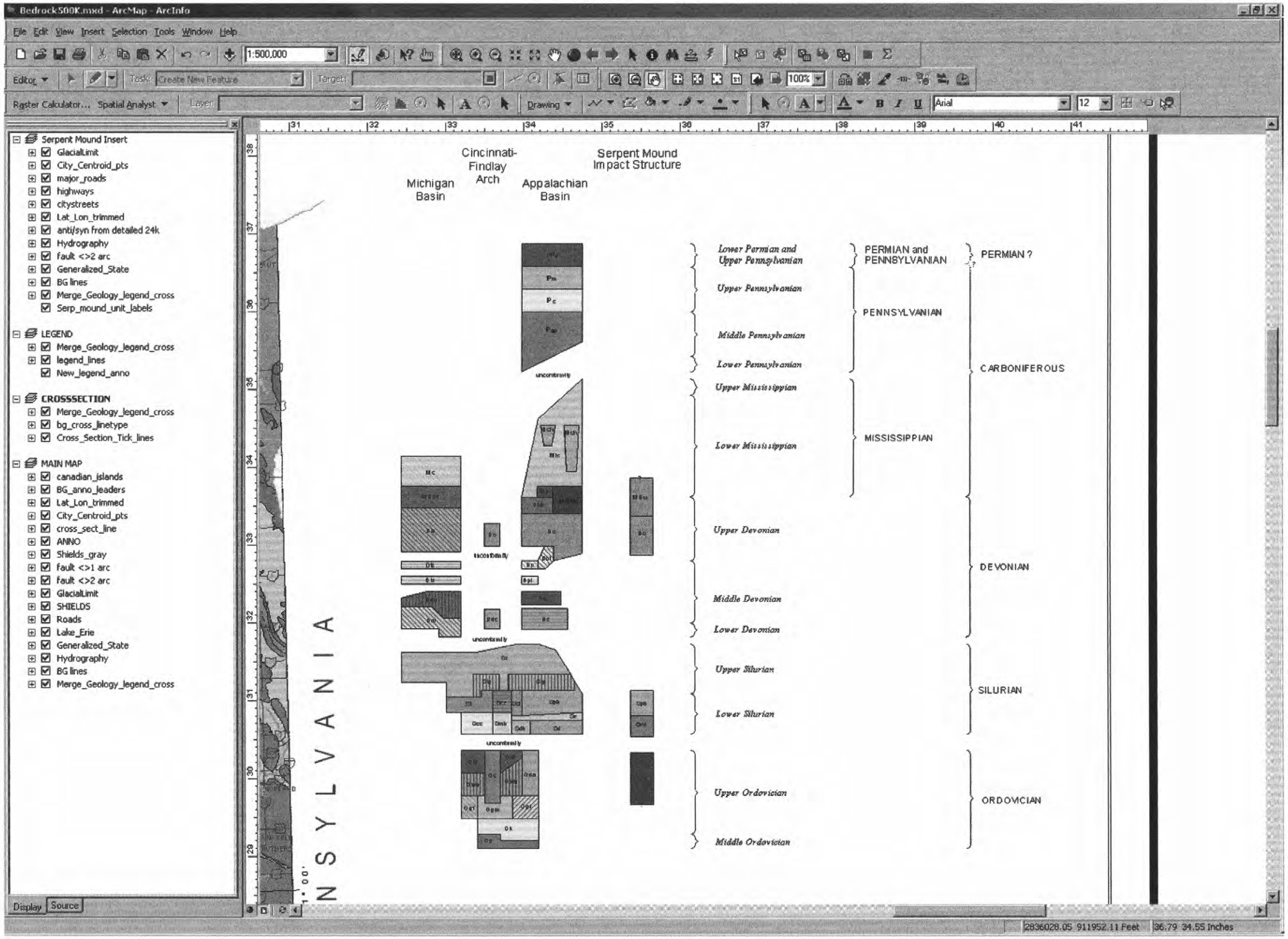


Figure 6. The cross section from the new state bedrock geology map is a GIS data layer.



Bedrock500K.mxd - ArcMap - ArcInfo

- Serpent Mound Insert
  - ☑ GlacialLimit
  - ☑ City_Centroid_pts
  - ☑ major_roads
  - ☑ highways
  - ☑ citystreets
  - ☑ Lat_Lon_trimmed
  - ☑ anti/syn from detailed 24k
  - ☑ Hydrography
  - ☑ fault <>2 arc
  - ☑ Generalized_State
  - ☑ BG lines
  - ☑ Merge_Geology_legend_cross
  - ☑ Serp_mound_unit_labels
- LEGEND
  - ☑ Merge_Geology_legend_cross
  - ☑ legend_lines
  - ☑ New_legend_anno
- CROSSSECTION
  - ☑ Merge_Geology_legend_cross
  - ☑ bg_cross_linetype
  - ☑ Cross_Section_Tick_lines
- MAIN MAP
  - ☑ canadian_islands
  - ☑ BG_anno_leaders
  - ☑ Lat_Lon_trimmed
  - ☑ City_Centroid_pts
  - ☑ cross_sect_line
  - ☑ ANNO
  - ☑ Shields_gray
  - ☑ fault <>1 arc
  - ☑ fault <>2 arc
  - ☑ GlacialLimit
  - ☑ SHIELDS
  - ☑ Roads
  - ☑ Lake_Erie
  - ☑ Generalized_State
  - ☑ Hydrography
  - ☑ BG lines
  - ☑ Merge_Geology_legend_cross

2836028.05 911952.11 Feet   36.79 34.55 Inches

Figure 7. The correlation chart from the new state bedrock geology map is a GIS data layer.

relief on the bedrock surface. These difficulties prevented the statewide usage of this data set; it was necessary to recompile the bedrock topography. The 1:24,000-scale bedrock topography contours from the GIS data set were used as the source for the newly compiled data layer. The contours were regridded, at a spacing of 60 meters, to produce a consistent bedrock topography GIS layer. To produce a bedrock topography surface across the entire state, the shaded surface-elevation grid of the state (Powers, Laine, and Pavey, 2002) was clipped to the edge of the bedrock topography grid. The clipped surface-elevation grid, which shows the unglaciated portion of the state in southern and southeastern Ohio, was then merged with the bedrock topography grid to produce the bedrock topography surface across the entire state (figure 8, Ohio Division of Geological Survey, 2003).

SOFTWARE APPLICATION

As the completion of the bedrock geology and bedrock topography GIS conversion projects neared, it became apparent that a methodology had to be developed for data management and data distribution of these large statewide GIS data sets. As part of a contract with ODOT for the completion of the GIS conversion, a multi-use application was created to help manage, distribute, and edit the data and maps. The application was designed with a number of different goals in mind. The application had to be able to extract and set up individual quadrangles for plotting and to extract digital data for public distribution for all 1,576 bedrock geology and bedrock topography maps. The application also had to enable dynamic updating and editing of the data so that

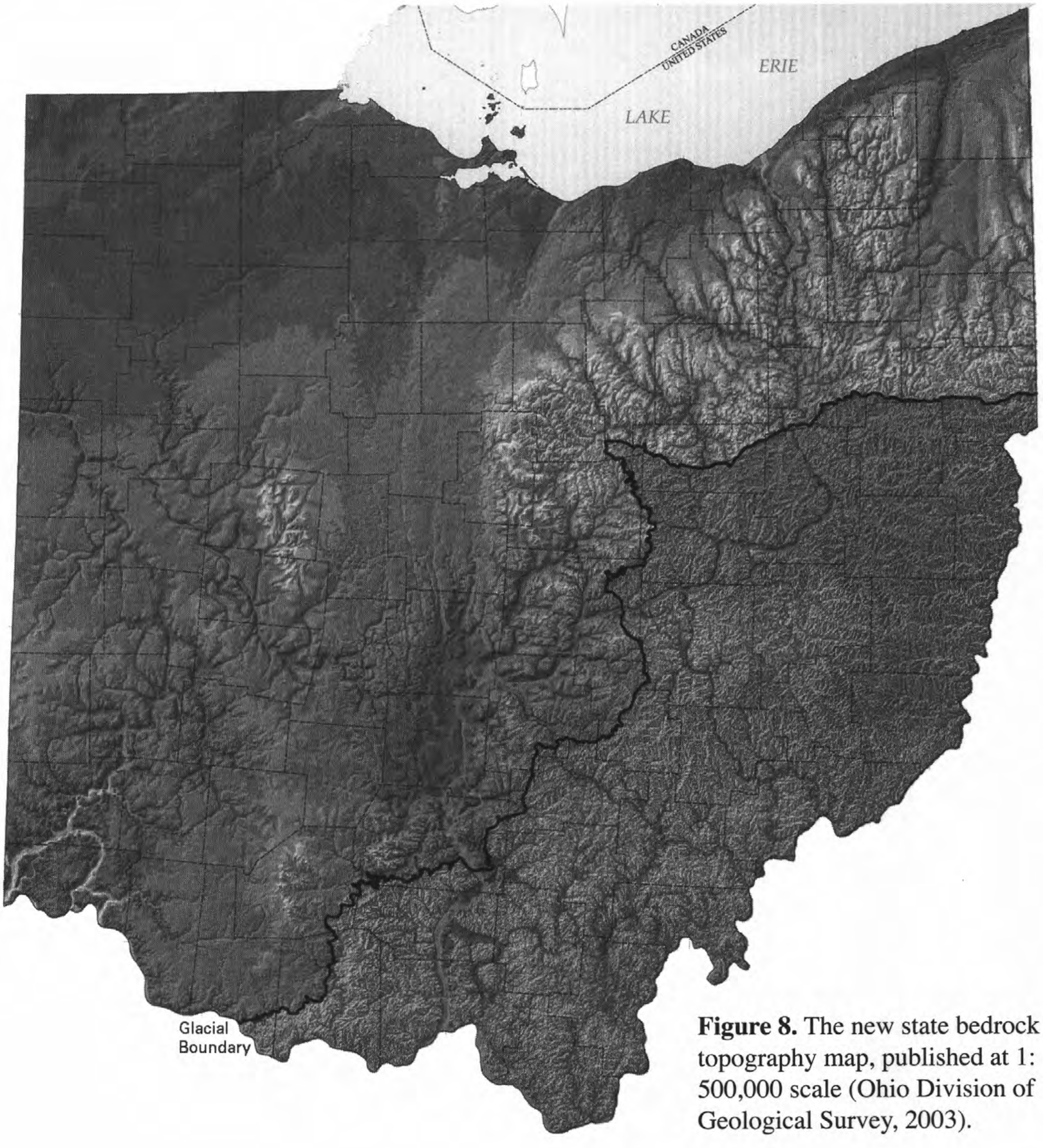


Figure 8. The new state bedrock topography map, published at 1: 500,000 scale (Ohio Division of Geological Survey, 2003).

versions released to the public are always current.

The application will assist the ODGS in managing and distributing maps and data to the public. One of the application's functions enables the user to select an individual 7.5-minute quadrangle either by drop-down list of quadrangle names or by selecting it from an index map of quadrangles covering the state of Ohio. Once selected, the bedrock geology or bedrock topography is displayed only for that selected quadrangle (figure 9). Also included are the bedrock topography contours, the bedrock topography data points, and the structure contour maps as TIFF images.

A second function allows for the export of the bedrock geology or bedrock topography data, along with the USGS Digital Raster Graphic base map. The bedrock geology or bedrock topography files are exported in shapefile format, and are written to a temporary directory. Once exported, the data can be distributed to the public.

A third function sets up the map layout for the traditional printing of bedrock geology and bedrock topography maps (figure 10). The layout function takes the extracted bedrock geology data layer, and exports it to a temporary shapefile. The temporary shapefile is then brought back into the map layout for standard symbolization and display. During the setup, the title, authors, and revision date are extracted from a quadrangle metadata database and placed onto the map layout. An inset map is automatically generated during the customized layout process, showing the location of the individual 7.5-minute quadrangle and the state outline. The legend is automatically generated, so that the geologic units will always have the same colors and patterns across the entire state. The legend for the geologic units uses a custom ESRI-style library that was developed for the application. During the quadrangle extraction process, the geologic unit in the attribute table is matched to the custom style in the ESRI-style library.

The fourth function is used only for plotting 1:100,000-scale bedrock geology and bedrock topography maps. This function has the ability to plot 30 x 60-minute blocks of map data that either correspond to a single USGS 1:100,000-scale topographic map, or to an area containing pieces of several USGS topographic maps.

The quadrangle-metadata database is used for multiple purposes. The title, authors, and revision date for the map layout come from the database. The database is used for bibliographic purposes and is used to update the National Geologic Map Database (NGMDB) Map Catalog. Future functions of the application may include placing a standard bibliographic citation on the maps, as obtained from the quadrangle metadata database. Another function may create individual metadata records, so when the map data is exported, customized metadata records can be exported with it.

SUMMARY

ODGS has created the 1:24,000-scale and 1:500,000-scale bedrock geology and bedrock topography GIS data sets, based upon a new set of 1:24,000-scale bedrock geology and bedrock topography maps produced by the ODGS. The resultant 1:500,000-scale bedrock geology map represents the first major revision of the official bedrock geology map of Ohio since 1920. An application has been created for plot-on-demand and data distribution of the bedrock geology and bedrock topography maps. A total of 1,576 individual 7.5-minute quadrangle maps can be extracted from the 1:24,000-scale bedrock geology and bedrock topography data layers. The maps can be set up for printing and the data can also be exported in a compressed ("zipped") shapefile format and distributed to the public. ODGS plans to extend the application for the extraction and plotting of 1,840 structure contour maps, 220 abandoned underground mine maps, and 120 karst geology maps that have been previously created or that are in preparation.

A quadrangle metadata database has been created to extract the titles, authors, and revision dates for the plotted maps, and to populate the NGMDB. Future implementations could include creating unique metadata records for all the extracted digital data and also creating unique bibliographic citations for all the plotted maps.

The new 1:500,000-scale bedrock geology map will be published in 2004. The new 1:500,000-scale bedrock topography map and the GIS files for the 1:24,000 and 1:500,000-scale bedrock topography recently have been released (Ohio Division of Geological Survey, 2003). Once the printed statewide bedrock geology map is published, the 1:24,000-scale plot-on-demand maps and GIS files for the 1:24,000-scale and 1:500,000-scale bedrock geology also will be released to the public.

REFERENCES

Bernhagen, R.J., 1994, A state geologist's memories of the Morrow County Boom, *in* Schafer, W.E., ed., Morrow County, Ohio "Oil Boom" 1961-1967 and the Cambro-Ordovician reservoir of central Ohio: Ohio Geological Society, p. 79-81.

Bownocker, J.A., 1920, Geologic map of Ohio: Ohio Division of Geological Survey Map 1, 1:500,000 scale, 1 sheet.

DeLong, R.M., 1965, Geology of the Kensington quadrangle, Ohio: Ohio Division of Geological Survey Report of Investigations 55, 1 sheet.

Ford, J.P., 1987, Glacial and surficial geology of Cuyahoga County, Ohio: Ohio Division of Geological Survey Report of Investigations 134, 29 p.

Larsen, G.E., 1991, Historical development and problems within the Pennsylvanian nomenclature of Ohio: Ohio Journal of Science, v. 91, p. 69-76.

McDonald, James, 2002, Computer-aided structure-contour

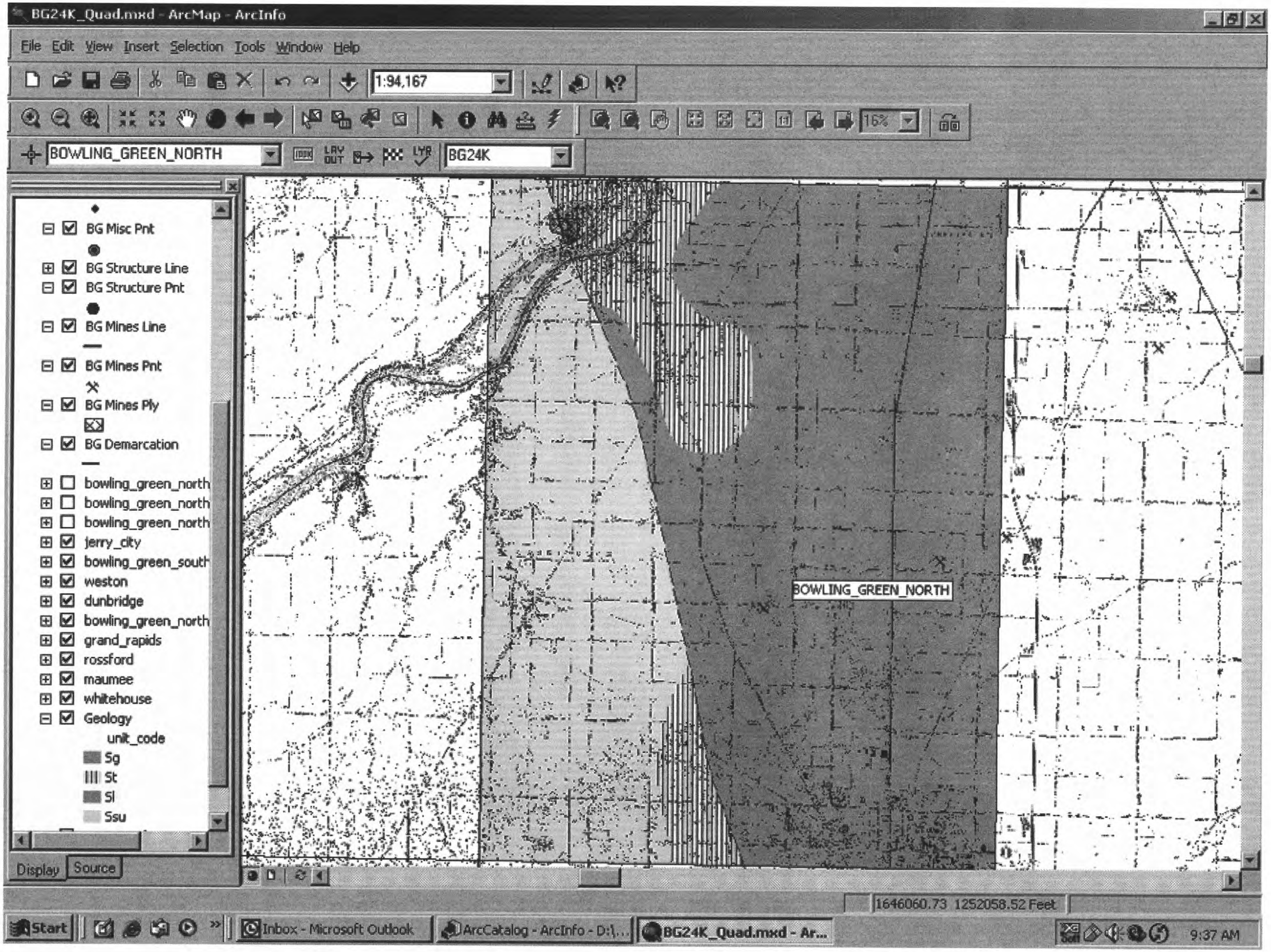


Figure 9. Extracted bedrock geology map of the Bowling Green North 7.5-minute quadrangle.

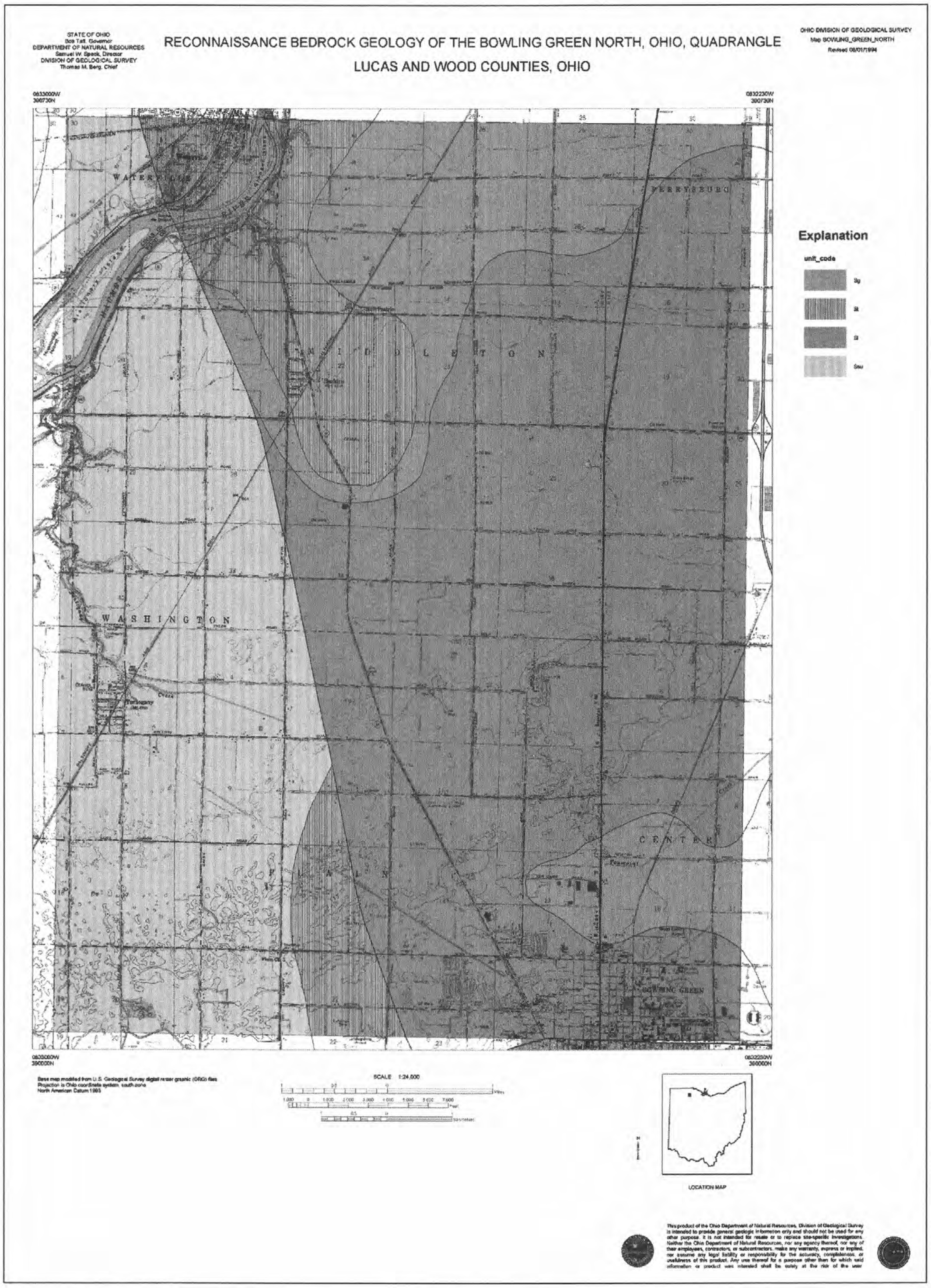


Figure 10. Bedrock geology map of the Bowling Green North 7.5-minute quadrangle, set up for plotting at 1:24,000 scale. The title and revision date on the map are extracted from a quadrangle-metadata database.

mapping in support of the Ohio Division of Geological Survey bedrock-geology program, in Soller, D.R., Digital Mapping Techniques '02—Workshop Proceedings: U.S. Geological Survey Open-File Report 02-370, p. 119-127, <http://pubs.usgs.gov/of/2002/of02-370/mcdonald.html>.

Ohio Division of Geological Survey, 2003, Bedrock-topography data for Ohio: Ohio Division of Geological Survey BG-3, 1 CD-ROM, GIS file format.

Powers, D.M., Laine, J.F., Pavey, R.R., 2002, Shaded elevation map of Ohio: Ohio Division of Geological Survey Map MG-1, 1:500,000 scale.

Sherman, C.E., 1933, Miscellaneous data: Ohio Cooperative Topographic Survey, Volume IV of the Final Report, 327 p.

Swinford, E.M., 1997, Remapping of the bedrock geology of Ohio completed: Ohio Division of Geological Survey, Ohio Geology, Summer 1997, p. 1, 3, <http://www.dnr.state.oh.us/geosurvey/oh_geol/97_Summer/remap.htm>.

Wickstrom, L.H., McDonald, James, and Berg, T.M., 1998, Working with the Department of Rehabilitation and Correction for the public good: Ohio Division of Geological Survey, Ohio Geology, Summer 1998, p. 1, 3, <http://www.dnr.state.oh.us/geosurvey/oh_geol/98_summer/leadart1.htm>.

Digital Karst Density Layer and Compilation of Mapped Karst Features in Pennsylvania

By Stuart O. Reese, P.G. and William E. Kochanov, P.G.

Pennsylvania Geological Survey
3240 Schoolhouse Road
Middletown, PA 17057
Telephone: (717) 702-2017
Fax: (717) 702-2065
e-mail: streese@state.pa.us, wkochanov@state.pa.us

INTRODUCTION

In 1985, the Pennsylvania Geological Survey began a series of investigations to map karst features of carbonate rocks in Pennsylvania (fig. 1). Locations of surface depressions, sinkholes, surface mines, and caves were compiled from municipal questionnaires, field surveys, published literature, and unpublished data sources as well as through an extensive aerial photograph review. Results of these investigations were released in a series of county-based open-file reports (Kochanov, 1987-1995). This information has been available since 1998 in paper maps and through an online database, but no digital layer files had been developed for geographic information system (GIS) tools.

Over the course of several months, a digital compilation of karst data points was completed. In total, the compilation included over 111,000 data points from 14 counties and 107 7.5-minute quadrangles (fig. 2). Digital data allowed for GIS mapping and for computer analysis of karst features. A colorized density surface was created from these merged files using ArcView 3.2 software. Mapping the density of karst points is useful for assess-

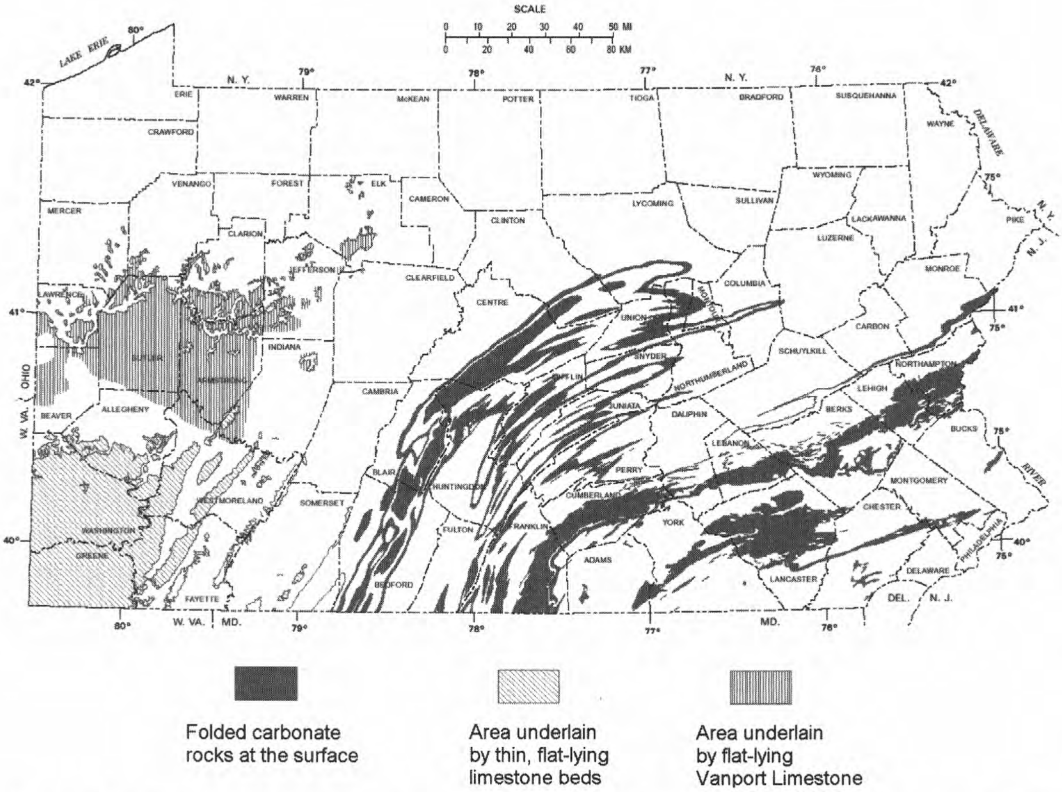


Figure 1. Carbonate rocks in Pennsylvania (modified after Pennsylvania Geological Survey, 2000).

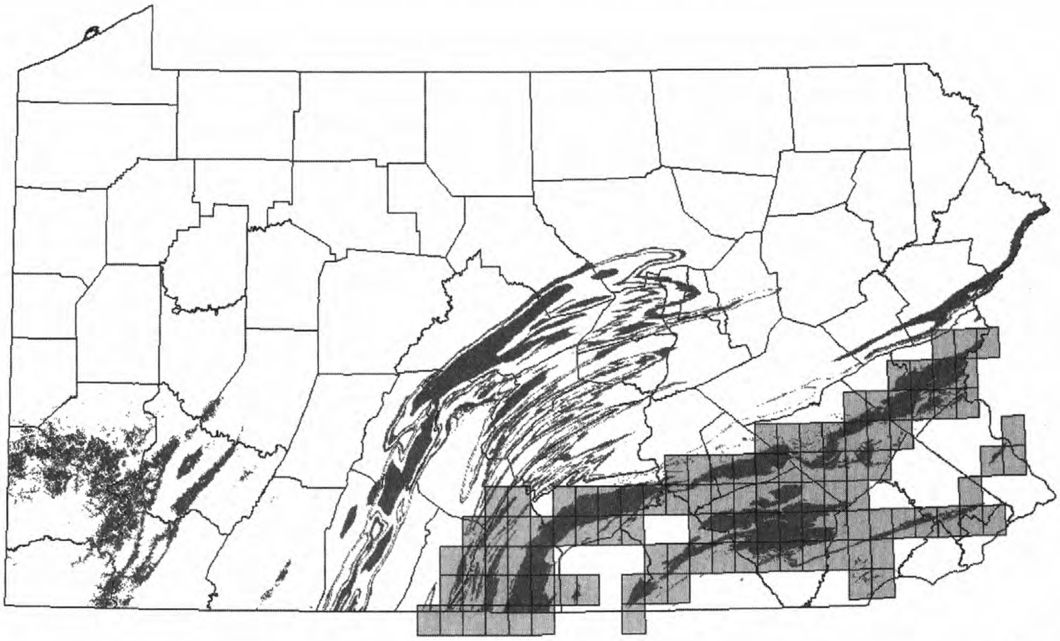


Figure 2. Location of quadrangles with mapped karst features in south-central and southeastern Pennsylvania.

ment of potential structural and environmental problems associated with karst geology. High-density areas of karst points where land subsidence may be a problem are noted, or where karst features can serve as direct recharge zones to the groundwater. These areas are highly vulnerable to groundwater contamination.

DIGITAL PROCEDURES

Data Compilation

Previously, karst feature locations from the open-file reports were plotted on a mylar stable base over the corresponding topographic map. The data points were digitized from the mylars using GSMAP v.8 (Selner and Taylor, 1992) and their coordinates were entered into a relational database management system. Karst data points were recorded in latitude/longitude(decimal degrees) by quadrangle. ArcView 3.2 was used to compile the karst point data from the database. ArcView shapefiles were created and merged to form a regional dataset. The shapefiles were placed over base maps of digital raster graphic images of 7.5-minute quadrangles in Universe Traverse Mercator (UTM), North American 1927 datum, Clarke 1866 spheroid, in UTM Zones 17 or 18.

The digital database was cross-checked against the original locations. As the data were reviewed, it became apparent that for many of the quadrangles, a systematic digitizing error had been introduced into the latitude and longitude data. These data points were corrected using the ArcView extension ShapeWarp (Version 2.2). In addition, on-screen digitizing procedures were used to create new files where previous digitizing had not been done. Compiled karst data points were identified by feature type (surface depression, sinkhole, surface mine, or cave), quadrangle, and county. County name was assigned by a spatial join command (ArcView geoprocessing, assign data by location).

Density Surface Preparation

The ArcView extension Spatial Analyst (version 2.0) was used to develop the digital density layer of the combined data points. A density surface is based on the division of the study area into square cells, which can be sized as appropriate. ArcView software calculates a density value for each cell by counting the number of points within a defined search radius from the center of each cell (fig. 3) and dividing by the search area. The density value (features per square mile) is assigned to the cell. The search circle is then shifted to the next cell and the floating process is repeated until all of the cells have been assigned a density value. This process smoothes the density layer over the study area.

ArcView has two options for density calculations – a simple density formula (described above) and a weighted "kernel" procedure. A weighted method called a kernel (a "quartic approximation to a Gaussian kernel") can be used which assigns more value to points located closer to the center of the cell. For this project, the simple density function was used. The kernel method further smoothes the data but in this case caused a more bi-modal appearance of the density surface. For this reason, the simple density surface was retained.

Density calculations must be done in a map projection that minimizes error for area, distance, and direc-

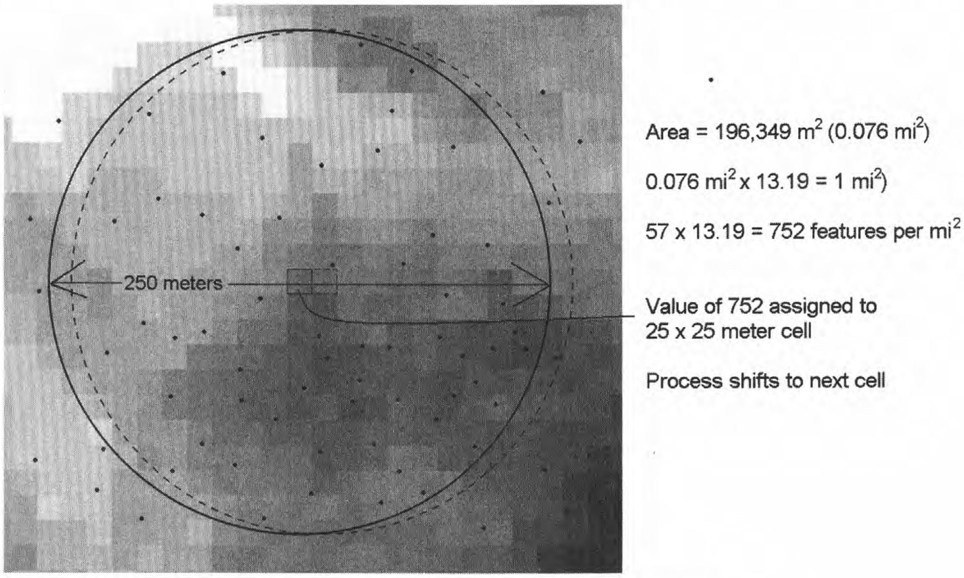


Figure 3. Procedures used in density calculation.

tion. For example, a density calculation using a decimal degrees (geographic) framework would result in severe errors. The projection used for the density calculations was the Albers equal area conic projection with standard parallels at 40°N and 42°N, and a central meridian of -78°W. This projection maintains true area and shape with negligible distortion at the scale used, which was less than or equal to 1:24,000.

When doing density calculations, the recommended number of cells is between 10 and 100 cells per density unit (Mitchell, 1999). The density unit of features per square mile was used here, which, using this recommendation, would equal approximately 100 cells (with dimensions of 160 x 160 meters) per square mile. This results in a 5 MB file size with over 1.3 million cells. At a scale of 1:24,000 (the scale at which the features were mapped), the density surface has a noticeably blocky appearance. A smaller cell size will smooth the surface, but require more computer processing time and file storage space. A cell size of 25 meters was chosen to smooth the data to the lower limit of mappability at the 1:24,000 scale. This produced a 213 MB file with over 55 million cells. Despite the larger file size, the resulting density surface portrays a smooth gradation of cells and allows more local variation to be seen at the 1:24,000 scale.

The chosen search radius influences the appearance of the density surface. The larger the radius, the more generalized the patterns will be. The smaller the radius, the more local variation is portrayed, to the extreme of remapping the point data. The 250-meter search radius was selected in a trial and error process to show enough local variation without over-generalizing the density. Karst feature data points were overlaid upon the density surface to evaluate different parameter values. The 250-meter radius parameter provides for a smooth simple calculated density surface.

Further preparation of the density surface was accomplished using ArcGis 8.3. The cells were color coded using a graduated color scheme for the density values. The density color scale that was developed represents an ESRI ArcMap 8.3 "quantile gradation" of the density values. Quantile values are useful in comparing the density over an area, and from map to map. The large number of classes (30 quantiles) allows the color gradation to be displayed as a range and it allows the value of "0" to be given a transparent definition on the map. The mapped areas of higher density of karst features are portrayed in orange and red colors. Values in the red approach 640 karst features per square mile (one feature per acre) or more. The lowest density value, represented by the darkest green color, indicates at least one karst feature within the 250-meter search radius of the cell (approximately 48 acres or 0.07 mi^2).

The units of measurement were selected to consider the proposed audience of the product. Although units of meters were used to develop the density grid, density was calculated for each grid cell in units of karst features per square mile (1 mi^2 = 2.59 km^2). A value in acres equivalent to the square mile value was added to the karst map. Units of square miles (and acres) allow the non-scientist to more easily relate to the map data. In the maps generated, yellow, orange and red colors approach 640 features per mile, or about one per acre.

DIGITAL PRODUCTS

The main products include digital coverages of karst points by county and by quadrangle, and a regional map showing the density of the karst features. Figure 4 shows an example of the density layer. County maps in

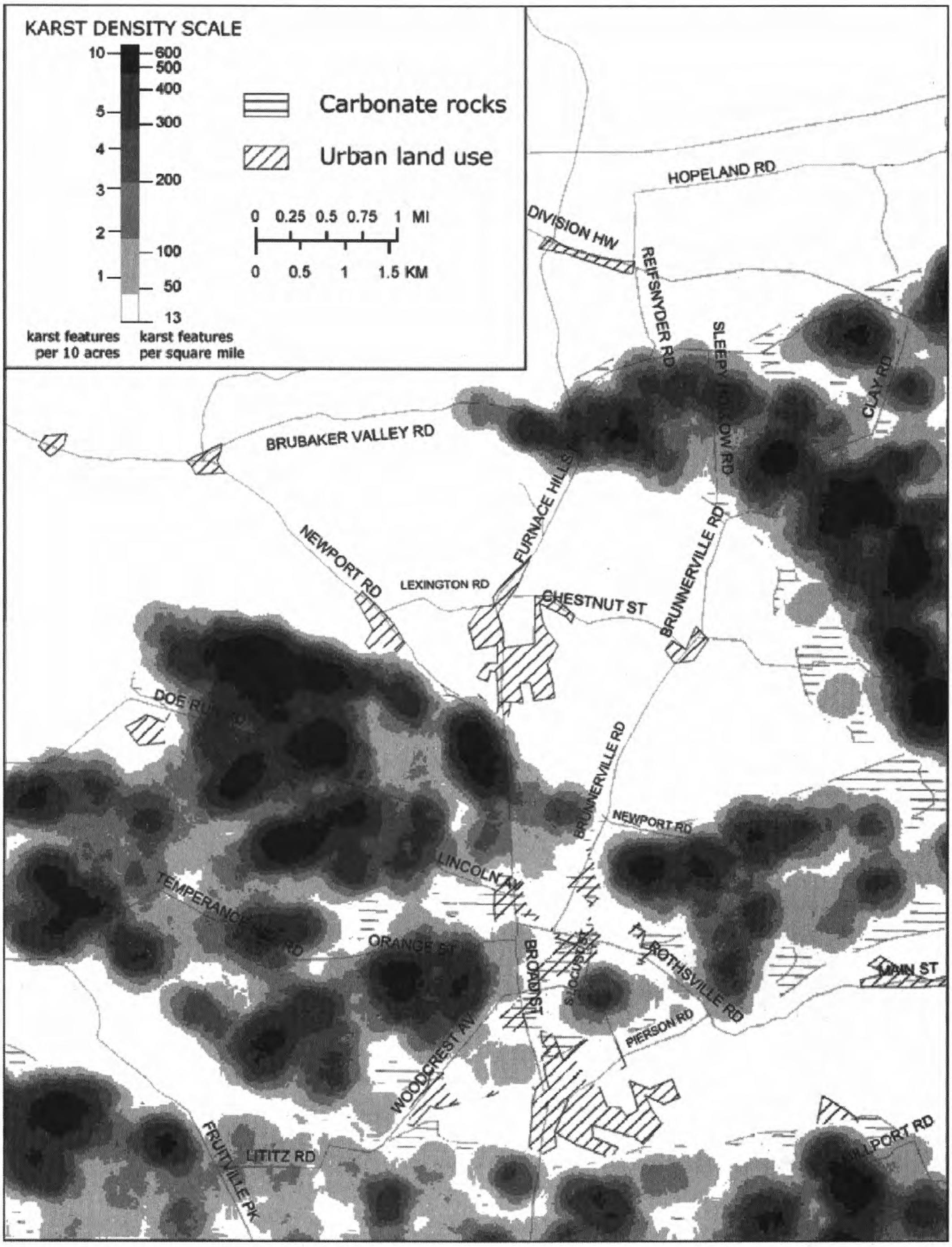


Figure 4. Density of mapped karst features in the Lititz 7.5-minute quadrangle, Lancaster County, Pa.

a mapbook style were completed. Such easy-to-interpret maps will help homeowners, municipal planners and others understand the intensity of mapped karst features per square mile or acre. Because of potential misinterpretation, efforts were made to qualify the digital products.

CAVEATS AND CONCLUSIONS

Beyond the accompaniment of metadata with a GIS coverage, explicit caveat statements are very important because of the multiple meanings that could be interpreted from the karst data. On the regional karst density map, the bright colors are quickly noticed as areas of concern. However, the proper response needs to be guided. Banning all activity in the high-density zones is not a rational response, but neither should caution be thrown to the wind. The recommended uses of the data should be made obvious. Here, the karst data are useful for regional planning and preliminary site studies, but they are not a substitute for site-specific subsurface investigations.

The occurrence of a sinkhole, a subsidence feature that breaks the land surface, depends on numerous factors including rock type, geologic structure (the presence of fractures, joints and faults), soil cover, surface hydrology, and land use. Areas of subsidence are not necessarily restricted to the high-density areas shown in dark shades on the map. Surface depressions, which by definition do not show a land-surface break, were the dominant type of mapped karst feature (96 percent). However, subsidence can occur in areas where there are no discernible surface depressions, or where sinkholes are not observed.

In addition, there may be instances where subsidence features are shown outside the mapped limits of carbonate bedrock. Surficial material such as colluvium typically conceals the actual contact between non-carbonate and carbonate bedrock formations. Undetected faults also can displace carbonate bedrock and account for subsidence features outside the current mapped formational contacts of carbonate rocks.

Land use can bias the detection of the karst feature. Urban land cover often masks karst features, making them difficult to detect. Sinkholes, though highly visible and often disruptive when they occur, are typically quickly filled or covered. Therefore, mapped karst features are most often under-represented in the urban setting. Urban land cover often accelerates the formation of sinkholes through changes to surface water drainage. Land thought to be free of sinkholes may suddenly develop karst subsidence features, especially where the surface hydrology has been altered. In more rural areas, karst features can be difficult to discern in wooded areas, whereas karst features in fields are more easily detected.

All of the potential biases must be considered when using the karst density maps and the digital data of the karst features. Caveat statements are needed to provide proper direction on the use and interpretation of the digital products. Because the mapped features are based on interpretation, cautionary statements are extremely important to direct the use of such digital products, especially when there are known limits to the mapping process. An understanding of the limits of the data is crucial to the responsible use of the data.

ACKNOWLEDGMENTS

We wish to thank staff members of the Pennsylvania Geological Survey Dr. Jon Inners and Michael Moore for critically reviewing the manuscript.

REFERENCES

Kochanov, W. E., 1987a, Sinkholes and karst-related features of Lehigh County, Pennsylvania: Pennsylvania Geological Survey, 4th ser., Open-File Report 87–01, scale 1:24,000, 19 p, 6 maps.

Kochanov, W. E., 1987b, Sinkholes and karst-related features of Northampton County, Pennsylvania: Pennsylvania Geological Survey, 4th ser., Open-File Report 87–02, scale 1:24,000, 24 p, 10 maps.

Kochanov, W. E., 1988a, Sinkholes and karst-related features of Berks County, Pennsylvania: Pennsylvania Geological Survey, 4th ser., Open-File Report 88–01, scale 1:24,000, 13 p., 16 maps.

Kochanov, W. E., 1988b, Sinkholes and karst-related features of Lebanon County, Pennsylvania: Pennsylvania Geological Survey, 4th ser., Open-File Report 88–02, scale 1:24,000, 11 p., 4 maps.

Kochanov, W. E., 1989a, Sinkholes and karst-related features of Cumberland County, Pennsylvania: Pennsylvania Geological Survey, 4th ser., Open-File Report 89–02, scale 1:24,000, 16 p., 9 maps.

Kochanov, W. E., 1989b, Sinkholes and karst-related features of Dauphin County, Pennsylvania: Pennsylvania Geological Survey, 4th ser., Open-File Report 89–01, scale 1:24,000, 8 p., 5 maps.

Kochanov, W. E., 1989c, Sinkholes and karst-related features of Franklin County, Pennsylvania: Pennsylvania Geological Survey, 4th ser., Open-File Report 89–03, scale 1:24,000, 15p., 20 maps.

Kochanov, W. E., 1990, Sinkholes and karst-related features of Lancaster County, Pennsylvania: Pennsylvania Geological Survey, 4th ser., Open-File Report 90–01, scale 1:24,000, 12 p., 18 maps.

Kochanov, W. E., 1993a, Sinkholes and karst-related features of Bucks County, Pennsylvania: Pennsylvania Geological Survey, 4th ser., Open-File Report 93–03, scale 1:24,000, 9 p., 4 maps.

Kochanov, W. E., 1993b, Sinkholes and karst-related features of Montgomery County, Pennsylvania: Pennsylvania Geological Survey, 4th ser., Open-File Report 93–02, scale 1:24,000, 7 p., 5 maps.

Kochanov, W. E., Lichtinger, J. F., and Becker, Mona, 1993, Sinkholes and karst-related features of Chester County, Pennsylvania: Pennsylvania Geological Survey, 4th ser., Open-File Report 93–01, scale 1:24,000, 9 p., 10 maps.

Kochanov, W. E., 1995, Sinkholes and karst-related features of York County, Pennsylvania: Pennsylvania Geological Survey, 4th ser., Open-File Report 95–06, scale 1:24,000, 9 p., 10 maps.

Kochanov, W. E., and Miller, Rachel, 1995, Sinkholes and karst-related features of Adams County, Pennsylvania: Pennsylvania Geological Survey, 4th ser., Open-File Report 95–05, scale 1:24,000, 9 p., 6 maps.

Mitchell, A., 1999, The ESRI Guide to GIS Analysis, Volume 1: Geographic Patterns & Relationships: Redlands, Ca, Environmental Systems Research Institute, Inc.

Pennsylvania Geological Survey, 2000, Limestone and dolomite distribution in Pennsylvania: Pennsylvania Geological Survey Map 15, scale 1:2,000,000. http://www.dcnr.state.pa.us/topogeo/maps/map15.pdf.

Pennsylvania Geological Survey, 2003, Density of Mapped Karst Features in South-Central and Southeastern Pennsylvania, by William E. Kochanov and Stuart O. Reese, 2003: Pennsylvania Geological Survey Map 68, scale 1:300,000. http://www.dcnr.state.pa.us/topogeo/map68.

Selner, G.I. and Taylor, R.B., 1992, GSMAP, GSMEDIT, GSMUTIL, GSPOST, GSDIG and other programs, version 8 for the IBM PC and compatible microcomputers to assist workers in the earth sciences: U.S. Geological Survey Open File Report 92-217.

Southern California Areal Mapping Project (SCAMP) and Multidimensional Databases

By Douglas M. Morton

U.S. Geological Survey
Department of Earth Science
University of California
Riverside CA 92521
Telephone: 909-276-6397
Fax: 909-276-6295
e-mail: scamp@usgs.gov

INTRODUCTION

The digital world has created a number of developments and opportunities for geologists, albeit allowing opportunities for abuses. Probably the most important impact to date is the routine production of publication-quality digital geologic maps. These maps can be produced and released relatively quickly, easily revised, and made inexpensively available on the web, thus keeping the maps current and serving map-using customers most effectively. However, for most geologists the greatest appeal of 'going digital' is the opportunity to solve complex geologic problems, because concomitant with the development and production of digital geologic maps has come the increasing ability to address geologic problems through quantitative spatial analysis. This type of analysis has evolved from simple intuitive overlays having two variables, to more sophisticated, nonintuitive multi-component analyses, such as factor analysis. With the advent of distributive parallel processing, it is now also possible to engage relatively inexpensively in uncompromised dynamic (numerical) modeling.

Two of the goals of the Southern California Areal Mapping Project (SCAMP) are to unravel the geologic history of southwestern California and to predict geologic events that may affect people and infrastructure in the region. Due to hazards resulting from the large number of major active faults, particular emphasis is placed upon the origin and evolution of the San Andreas Fault System and attendant structural and geomorphic changes such as mountain and sedimentary basin formation. Other goals include landslide (especially debris flow) hazards analyses, and unraveling Cretaceous orogenesis and magmatism. Common to these divergent geologic goals is the requirement for high quality multidimensional areal and temporal databases. The most fundamental database is the two-dimensional digital geologic map with temporal and other attributes.

SOME EXAMPLES

The following briefly summarizes examples of the current status of (1) searchable geologic maps, (2) unraveling past processes; and (3) predicting events.

Searchable Geologic Maps

Development of a prototype searchable geologic map for southwestern California is currently underway, and is being coordinated with the National Geologic Map Database Project. The prototype area is a revision of a 60'x60' geologic map coverage of the recently released Santa Ana (Morton, 1999) and San Bernardino (Morton and Miller, 2003) 30'x60' quadrangles. This map covers a rapidly growing area that currently has a population of over 5 million, greater than the populations of 30 states. To produce the map, most of the polygon data are being input at 1:24,000 scale, about 15 percent at 1:12,000, and 5 percent at 1:62,500. Detailed geologic maps of selected areas at scales as large 1:600 and 1:1,200 will augment the 1:100,000-scale map. The geology within the area is complex, and includes over 675 map units, parts of 3 major geologic provinces of California (Mojave Desert, Transverse Ranges, and Peninsular Ranges), and several major active faults, including the San Andreas, San Jacinto, Elsinore, Whittier, and Cucamonga Faults. Due to a combination of steep slopes, the fractured nature of much of the bedrock, and the character of many Tertiary and Quaternary units, the area abounds with landslides. We are attempting to construct the map to the quality of USGS Geologic Quadrangle maps and Professional Papers, and will include interactive and searchable databases. Publication will have a standard Correlation of Map Units and Description of Map Units as well as extensive illustrated text describing the geologic history and features of the area.

SCAMP has closely interacted with the very extensive southern California map user community to determine what map attributes are needed to answer the most important and anticipated questions. Most user needs and queries to the database require combination of polygon and line attributes or polygon and point attributes. Also, users expressed considerable interest in relating recency of fault displacement to map units, especially Holocene deposits cut by faults. Currently only 20-25 attributes are being entered into spreadsheets for each map unit; these include age, name, and basic descriptions (mineral composition, texture, etc). Future revisions will contain additional attributes, but because there are 675 map units, this initial effort in itself is not trivial. It is anticipated that when a legend parser is completed in the near future by the National Geologic Map Database Project, data entry will be greatly facilitated.

To impart as concise a picture as possible of the geology, the map content builds on the adage that 'a picture is worth a thousand words.' The map will be augmented with digital images for most of the major map units at scales ranging from landscape to microscopic, as well as images of structural and physiographic features, particularly faults and landslides. The recently released San Bernardino 30'x 60' quadrangle includes 149 color photographs, less than 10% of what is planned for the Santa Ana-San Bernardino map. The exhaustive retinue of images will permit virtual geologic field trips through the area. In addition to digital images, analytical data such as major and trace element chemistry, isotope geochemistry, specific gravity, and magnetic susceptibility will be accessible by map unit, individual map polygons, and individual points within polygons.

Unraveling the past – Cretaceous orogenesis and magmatism

Digital geologic maps showing rock type distribution and structure are fundamental to understanding the Cretaceous history. About 60 million years(Ma) of geologic history of the Peninsular Ranges Province is recorded only in Mesozoic and older basement rocks. For the northern part of the province a number of databases were developed to help unravel the geologic setting and history for this interval of time. Analyses of the collective databases indicate that during the Cretaceous this part of the Peninsular Ranges Province was the site of a major and prolonged period of orogenesis that produced a wide and complex assemblage of volcanic, plutonic, and metamorphic rocks. Digital geologic maps showing rock type distribution and structure are fundamental to understanding the Cretaceous history. Extensive density, magnetic susceptibility, and major and trace element databases were developed using approximately 330 samples collected by the late A.K. Baird and his coworkers (Baird and others, 1979; Baird and Miesch, 1984) augmented by 200 new samples. In addition, Sr and Rb isotopic values were obtained for these samples and initial Sr values, (Sri) were calculated. The Sri values were used to characterize the source of the magma (Kistler and Morton, 1994; Morton and Kistler, 1997). For selected samples whole rock ^{18}O and common Pb isotopic values were determined (Kistler and others, 2003) to further understand the nature of magma source materials. Samples from a geochemical traverse across the northern part of the Peninsular Ranges batholith were analyzed for whole rock elemental and isotopic chemistry including Nd/Sm (Premo and others, 1994). Emplacement ages of plutonic rocks are based on zircon and sphene U/Pb ages, and cooling temperature ages are based on Ar/Ar ages from hornblende, biotite, and K-feldspar (Premo and others, 1994). Pressure of crystallization for hornblende was determined for some samples. Structural analysis, metamorphic age, temperature and pressure, and provenance of detrital zircons were determined for selected metamorphic rocks (Johnson and others, 2000; Premo and others, 2002). All of these data are integrated in the geologic map database, and are queryable.

Interpretations derived from the collective databases have unraveled parts of the Cretaceous history of this part of southern California. Emplacement ages of plutons indicate the batholith was developed over a 45 Ma interval. Magmatism, preceded by tectonism, began in the western part of the Peninsular Ranges at 125 Ma before present(BP), accompanied by volcanism and emplacement of plutons into Mesozoic marine sedimentary rocks that are probably part of a back-arc basin assemblage. Magmatism progressively extended eastward where plutons were emplaced into Paleozoic and Proterozoic (?) continental nonmarine rocks; magmatism continued to about 80 Ma(BP).

Based on extensive Sri and limited Nd/Sm data, magma that formed the plutons in the western part of the batholith was derived from oceanic crust. These plutons have relatively uniform Sri values, mostly between 0.703-0.704, have an extensive compositional range from gabbro to granite, and cover a remarkably broad area. The central part of the magmatism had a source from a mixed oceanic and continental crust; the eastern part has Sri values ranging from 0.707-0.708+, indicating a continental crust source. Plutons in the western part were passively emplaced, crystallized at pressures of 2-4 Kb, and are relatively small to intermediate in size and exposed at shallow levels. Similar sized plutons in the central part of the batholith were emplaced at intermediate to deep levels, and are relatively high-strain. They were in part forcefully emplaced and crystallized at pressures of 5-6.5 Kb. Further east, large plutons were emplaced about 100 Ma(BP), crystallized at lower pressures of about 4-4.5 Kb, and were accompanied by development of a major dislocation zone, producing a broad zone of Buchan thermal metamorphism and repeated deformation of the metamorphic rocks (Johnson and others, 2000; Morton and others, 2000; Bern and others, 2002). East of the large plutons is a regional mylonite zone, the Eastern Peninsular Ranges Mylonite Zone. Plutons east of the mylonite zone crystallized at high pressures of 6-6.5 Kb at 80-85 Ma(BP).

Prediction – El Nino soil slips-debris flows

Landslides of a wide variety abound in southern California. They constitute one of the most serious geologic hazards in the region and are of major interest to a large segment of our geologic map users. Debris flows are a common and widespread landslide type that occur by the tens of thousands during unusually wet "El Nino" winters. These landslides occur during periods of intense rainfall, beginning as soil slips - small slab-like failures that disintegrate to form debris flows that move various distances down slope. Although most are small in size, these flows can do considerable damage and result in loss of life. To mitigate damage and loss of life, it is important to be able to predict the size and the dynamics of future debris flows and when and where they will occur.

Over 100,000 debris flows have been mapped and systematically digitized, producing an essential database that can be used to help develop predictive tools for the occurrence of soil slips and debris flows. Debris flows that were generated under different rainfall conditions were mapped in 15 different geologic-physiographic settings in southwestern California. Debris flow maps were produced for a number of winters over the period of 1927 to 2001, with most of the debris flow data obtained for the years 1969 and 1998.

Based on analyses of selected variables, geology, slope, and aspect were the most useful for the production of predictive maps showing the locations having the highest likelihood to generate soil slips (Hauser, 2000; Koukladas, 1999). Over 700 geologic map units were assigned a numerical value based on the number of soil slips in the mapped units, per unit area. Similarly, numerical values were assigned to slope and aspect categories, again based upon the frequency of soil slips by slope and aspect. An algorithm for predicting the point of origin of soil slips was then developed from the geology, slope, and aspect values. For map presentation, soil slip susceptibility values were calculated from 5-meter cells of geologic map units; slope and aspect were calculated from 10-meter digital elevation models(DEMs). The resultant soil slip susceptibility values were divided into four categories, ranging from no susceptibility to low, medium, and high susceptibility. A susceptibility category was assigned to individual 10-meter cells. Soil slip susceptibility values were calculated for over 2,000,000 10-meter cells covering 128 7.5' quadrangles in southwestern California (Morton, Alvarez, and Campbell, 2003).

To test the accuracy of the soil slip susceptibility values three test areas, one coastal, one inland, and one semi-arid, were selected to compare actual mapped soil slip-debris flows with the soil slip susceptibility maps. The test showed that 85% to 95% of the soil slips occurred in high susceptibility value areas. Without digital maps these types of analyses are not possible. Work is continuing to refine the susceptibility maps and provide answers to timing, size, and dynamics of the debris flows.

FUTURE – DYNAMIC MODELING

The principal long range goal of SCAMP is to produce a dynamic model for the complex geologic history of southern California, and to predict future geologic events ranging from tectonism to landsliding. Included in the dynamic model will be interaction of tectonism, denudation, mass wasting, erosion, and sedimentation. There are now more comprehensive, faster, and less expensive computational means, utilizing widespread distributive computing such as Beowulf clusters, available to solve extremely complex problems. In order to utilize the staggering increases in computational ability the geosciences community faces vast challenges. At no time in the past was there ever the demand for 4-dimensional data that there is today, and that demand will be ever greater in the near future. Concomitant with development of new analytical techniques to further our insights on composition, temperature, pressure, and age of earth materials there will be greater demands than ever before on the field geologist to collect, in digital form, more detailed and comprehensive field data. The challenges of the future work load are staggering, but the prospects of fundamental new geologic insights unraveling the past and predicting the future are boundless.

REFERENCES

Baird, A.K., Baird, K.W., and Welday, E.E., 1979, Batholithic rocks of the northern Peninsular and Transverse Ranges, southern California: Chemical composition and variation; *in* Abbott, R.L., and Todd, V.R., eds., Mesozoic crystalline rocks: Dept. Geological Sciences, San Diego State University, San Diego, California, p. 111-132.

Baird, A.K., Miesch, A. T., 1984, Batholithic rocks of southern California – A model for the petrochemical nature of their source materials: U.S. Geological Survey Professional Paper 1284, 42 p.

Bern, A., Hammarstrom, J., Morton, D.M., Premo, W., and Snee, L.W., 2002, Plutonic and metamorphic rocks, northern Peninsular Ranges batholith, southern California – Structural and uplift history based on new geobarometric and isotopic data: Geological Society of America Abstracts with Programs, vol. 34, p. A-124.

Hauser, R.M., 2000, Soil slip-debris flows during the winters of 1926-27, 1968, and 1997-98 in the Santa Paula area, Ventura County, California: Riverside, California, University of California, M.S. thesis, 64 p.

Johnson, A.M., Snee, L.W., Morton, D.M., Hammarstrom, J., Miggins, D., Premo, W., and Yeoman, R., 2000, Structural and mineralogical changes across a Buchan metamorphic gradient, northern Peninsular Ranges batholith, southern California: P-T-T and tectonic implications: Geological Society of America Abstracts with Programs, vol. 32, p. A-168.

Koukladas, Catherine, 1999, Lithologic and slope aspect controls on infinite slope failures in the western San Gabriel Mountains, Los Angeles County, southern California: Reno, Nevada, University of Nevada, M.S. thesis, 199 p.

Kistler, R.W., and Morton, D.M., 1994, Sr, Rb, Sri variation and whole-rock Rb-Sr ages of plutons in the northern Peninsular Ranges batholith, southern California: Geological Society of America Abstracts with Programs, vol. 26, no. 2, p. 82.

Kistler, R.W., Wooden, J.L., and Morton, D.M., 2003, Isotopes and ages in the northern Peninsular Ranges batholith, southern California: U.S. Geological Survey Open-File Report. 03-489, 45 p.

Morton, D.M., 1999, Preliminary digital geologic map of the Santa Ana 30'x60' quadrangle, southern California, version 1.0: U.S. Geological Survey Open-File Report 99-172, <http://geopubs.wr.usgs.gov/open-file/of99-172/>.

Morton, D.M., and Kistler, R.W., 1997, Sri variation in the Peninsular Ranges batholith: Geological Society of America Abstracts with Programs, vol. 29, no. 6, p. A-69.

Morton, D.M., Premo, W.R., Kistler, R.W., Snee, L.W., and Hammarstrom, J.M., 2000, Structural and mineralogical changes across a Buchan metamorphic gradient, northern Peninsular Ranges batholith, southern California: Setting and structure: Geological Society of America Abstracts with Programs, vol. 32, p. A-168.

Morton, D. M., Alvarez, R.M., and Campbell, R.H., 2003, Preliminary soil-slip susceptibility maps, southwestern California: U.S. Geological Survey Open-File Report 03-17, <http://geopubs.wr.usgs.gov/open-file/of03-17/>.

Morton, D.M., and Miller, F.K., 2003, Preliminary geologic map of the San Bernardino 30'x60' quadrangle, California: U.S. Geological Survey Open-File Report 03-293, <http://geopubs.wr.usgs.gov/open-file/of03-293/>.

Premo, W.R., Morton, D.M., Kistler, R.W., Snee, L.W., Lichte, F.E., and Naeser, N., 1994, Preliminary geochemical and isotopic results of a section across the northern Peninsular Ranges batholith: Geological Society of America Abstracts with Programs, vol. 26, no. 2, p. 82.

Premo, W.R., Morton, D.M., Snee, L.W., and Bern, A.M., 2002, Shrimp U-Pb ages of provenance from detrital zircon populations of intra-batholithic metasedimentary rocks,northern Peninsular Ranges batholith, southern California: Geological Society of America Abstracts with Programs, vol. 34, p. A-124.

Digital Geologic Mapping in a Data Rich, Urban Environment

By Gayle H. "Scott" McColloch and Jane S. McColloch

West Virginia Geological and Economic Survey
P.O. Box 879
Morgantown, WV 26507-0879
Telephone: (304) 594-2331
Fax: (304) 594-2575
e-mail: mccolloch@geosrv.wvnet.edu; janemc@geosrv.wvnet.edu

INTRODUCTION

Is it possible to use data from GIS databases designed to generate tax bills to produce geologic maps? This is the question that we posed before beginning a pilot project in 2002, to map the West Virginia portion of the Lake Lynn 7.5-minute quadrangle and to explore utilizing pre-existing GIS data to produce geologic maps.

HISTORY OF DATA COLLECTION IN THE PENNSYLVANIAN OF WEST VIRGINIA

The West Virginia Geological and Economic Survey (WVGES) was established in 1897. Its initial tasks were to produce summary volumes on various resources or areas of interest and to produce a series of "county reports" with accompanying geologic maps. The commonly used term, "county report," is a misnomer; many include more than one West Virginia county, although the counties are always adjacent and geologically closely related. The fieldwork for each report was completed in a single field season. By modern standards the fieldwork appears to have been rushed. Errors of interpretation were, understandably, fairly common in geologically complex areas. Although some types of exposures like small underground mines and railroad cuts were common, others, notably surface mines and large highway road cuts, did not exist. The fieldwork produced much usable data, most of which was published. County reports were published between 1907 and 1939, with a significant hiatus during the Depression years when funding was more limited than usual.

The second phase of agency history, from 1939 through approximately 1970, involved studies such as "Germanium in Coals of West Virginia" or "Geology and Economic Resources of the Ohio River Valley in West Virginia" that focused on more specific themes and areas than the previously mentioned county report series. During this period, significant work by Thomas Arkle and Dr. Aureal T. Cross was conducted in the Pennsylvanian rocks of West Virginia. In addition, A. J. W. Headlee, John P. Nolting, Jr., Homer A. Hoskins, Ralph E. McClelland and Richard G. Hunter accomplished much coal analytical work. The works of these individuals contributed to the knowledge base provided by the county reports (West Virginia Geological Survey, 1963). During this period, the county report maps were generalized to produce an up-to-date 1:250,000 geologic map.

The next phase of Pennsylvanian studies at the WVGES began in the early 1970's, when the legislature appropriated funds to begin a new data collection effort in the Pennsylvanian coal measures. The resulting project was called the "Coal Resources and Pollution Potential Study." Funding for this study was gradually reduced in the early 1980s and employees left the agency or were reassigned to other work. Funding to publish results was not provided. Although few formal publications resulted, this study produced a large body of carefully organized raw data and useful working maps, which were given open-file report status. This collection has been extensively used for many purposes over the past 25 years.

The final chapter of Pennsylvanian data collection and publication began in 1995. A long-term controversy concerning how West Virginia mineral resources were taxed erupted into the courts. The settlement made funding available for a new appraisal methodology, and the WVGES was charged with project management and development of GIS-based models of mineral resources. The West Virginia Division of Tax and Revenue was responsible for developing a statewide GIS-based mineral property database. Funding was also provided for a private consultant to develop appraisal methodology and software, to establish a GIS Technical Center at West Virginia University, and to hire a state GIS Coordinator. The resulting project, including all participants, was termed the Mineral Lands Mapping Project (MLMP).

As the WVGES has been developing a georeferenced database of oil and gas well information since the 1960s, the major part of the MLMP at WVGES involves the compilation of all available coal information into GIS databases. This effort is termed the Coal Bed Mapping Program (CBMP). As a rich body of raw field data is available from the efforts of the earlier Coal Resources and Pollution Potential Study and core logs provided by the coal industry, the CBMP involves the completion of data collection from tens of thousands of mine maps and the compilation of all of this data into a large statewide GIS data set. Major products of CBMP include the following 1:24,000 coverages and grids for each major coal seam:

- Bed structure (vector contours and grid)
- Bed thickness (vector contours and grid)
- Bed percentage partings (vector contours and grid)
- Mining
- Bed discontinuities
- Coal boundaries
- Thickness point locations
- Elevation point locations

Although primarily produced to characterize mineral resources, several of these coverages and grids are useful for geologic mapping. The mining layer is also useful for environmental applications.

GEOLOGIC MAPPING

Although coal beds are considered "informal units" in the North American Stratigraphic Code (The North American Commission on Stratigraphic Nomenclature, 1983), they are frequently the most continuous, traceable units in the Pennsylvanian rocks of the Appalachian Basin. For this reason they are frequently used as rock unit boundaries. Geologic mapping utilizing CBMP data is possible because a few major coal beds form the boundaries of some mapping units in the region (fig. 1). The boundaries of pre-Pennsylvanian non-coal bearing units can be obtained by creating structure contour maps of various horizons using georeferenced oil and gas well data.

The CBMP product includes two carefully developed GIS data sets: a polygon coverage representing the extent of each important coal bed, and elevation grids representing the structure of each coal bed. The grids are divided into quadrangle-sized areas for convenience in processing. Boundaries of polygons that represent coal bed outcrops are carefully determined; however, the boundaries of polygons that represent the extent of coal beds below drainage or where no mining has occurred are somewhat subjective. The procedure for generating coal bed outcrops involves intersecting the coal bed structure grids and the grids representing the elevations of the surface by subtracting the value of the surface elevation from the elevation of the coal beds and contouring the zero value. This single contour line represents an approximate version of the outcrop (fig. 2). This outcrop line is then edited by overlaying it on a carefully georeferenced image of the 7.5-minute topographic map to incorporate indications of the outcrop position such as underground mine openings, surface mines, and breaks in slope.

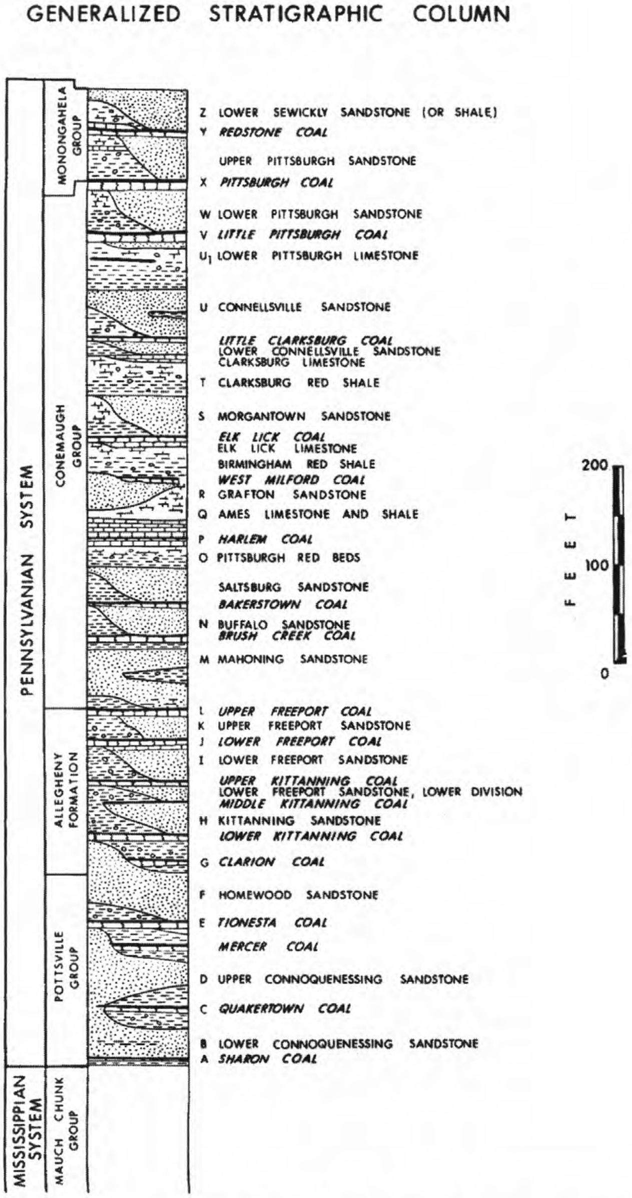


Figure 1. Section from Fonner and others, 1981, illustrating coal beds bounding mapping units.

The Northern Appalachian Coalfield is known to have consistent sedimentary thicknesses. This regularity is not only true of the Pennsylvanian, but also of older units occurring near the surface. This allows extrapolation of important contacts within stratigraphic units. The important basin-wide unconformity between the Harlem coal bed and the Ames Limestone, which represents the last marine incursion into the Appalachian Basin, is mappable{?} on the Lake Lynn 7.5-minute quadrangle. For mapping purposes, this contact was generated by determining

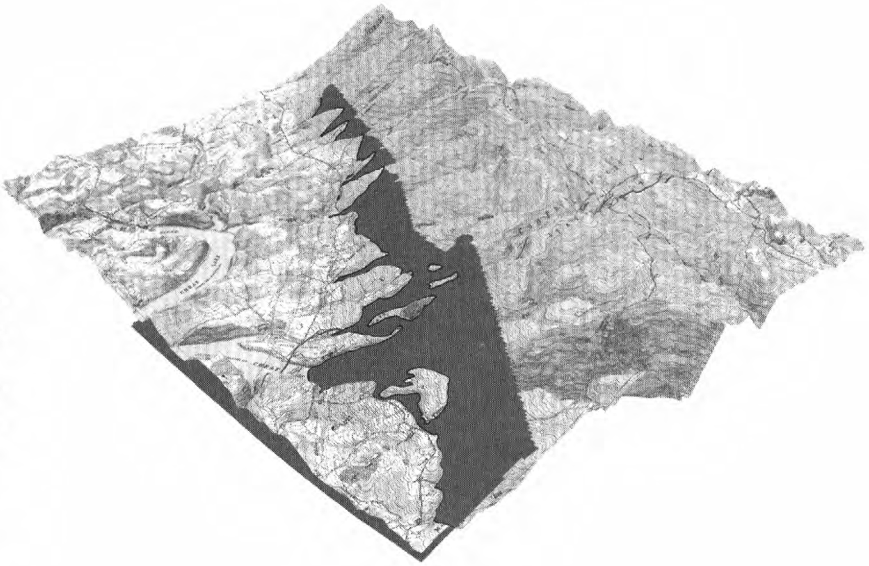


Figure 2. Image of Lake Lynn 7.5-minute topographic quadrangle draped over an elevation model of the quadrangle. The gray horizon represents the base of the Upper Freeport coal. The black row of pixels at the intersection represents an approximate outcrop defined by contouring the intersection of the coal and the land surfaces.

the interval between the Harlem/Ames contact where it was well exposed in a highway road cut on an adjacent quadrangle and the underlying Upper Freeport coal (see fig. 1). An outcrop line for this contact was extrapolated using the elevation grid of the Upper Freeport coal. The placement of this contact was verified by comparison of features on the Lake Lynn topographic map and field observations.

This comparison showed that the procedure described above to map the unconformity between the Harlem coal and the Ames Shale allowed correlation of the coal mined in a surface mine on the quadrangle and positive identification of an outcrop exposed when the level of Cheat Lake was dropped for sewer line construction in the spring of 2003. The location of the Pittsburgh and Upper Freeport coal outcrops was correctly delineated on the CBMP dataset with the exception of one area where the elevation of the Upper Freeport coal had to be corrected to agree with field observations.

URBAN MAPPING

Although West Virginia is well known as one of the least urbanized states in the United States, it does have some urban areas. One of the latest to be formally recognized, as a result of the 2000 census, is the Morgantown Metropolitan Statistical Area, which includes Monongalia and Preston counties. This designation has generated the need for up-to-date geologic maps that incorporate all available data.

Some unique issues need to be addressed when collecting data and producing maps in urban areas. The first is access. Greater employment opportunities are drawing people to the metropolitan areas. Two-income households are becoming more common in these areas of West Virginia, and an increasing number of property owners are not at home during the day. Near the boundaries of the Morgantown urban area, relatively small properties display a profusion of "No Trespassing" signs. Unlike the farmers of more rural areas of the state, the residents are frequently absent when a mapper wants to ask permission for access. Most do not understand why a geologist wants to look at the rock outcrop in their backyard. The rich body of pre-existing data, including several generations of air photos and other aerial imagery, is useful because field observations can frequently be made from public roads and other accessible areas.

The second issue is locating unit contacts as accurately as possible. One instance illustrates this point. The WVGES received a call from a consultant working in MacAurthur, West Virginia, near Beckley, in the early 1980s. The initial question concerned the interval between the Sewell and the Beckley coals. While digging a ditch for a new sewer line, a backhoe had cut into old mine works. Apparently a preliminary WVGES map showed the outcrop of the Sewell coal was located on the hillside above where the ditch was being dug, and the consultant assumed the old mine works were in the stratigraphically lower Beckley coal. At this location the interval between the Beckley and Sewell coals is about 285 feet. The difference in elevation between the Sewell coal contact on the map and that of the old mine works was less than 40 feet, which ruled out the old mine works being in the Beckley coal. The mapped location of the outcrop was slightly inaccurate, and the seam in question was undoubtedly the Sewell.

Soon afterward, one of the authors drove by the site and found a mine subsidence control project located at a nearby West Virginia Department of Natural Resources, Abandoned Mined Lands Office. That agency was desperately attempting to save the buildings, which were adjacent to the sewer line ditch (fig. 3). The buildings predated the preliminary map, but had the subsiding buildings been sited using information from the map, the WVGES might have been held responsible. This experience and many others have led to a realization of how critical the location of a geologic contact or outcrop can be, especially in areas where the population is increasing.

The scale problem associated with some symbols used on topographic maps is another urban mapping issue that could adversely affect the quality of an urban geologic map. Most geologists are aware that several classes of symbols used on topographic maps are not printed to scale. The scale of symbols is not very important on maps of rural areas where adequate space exists so that oversized symbols can be plotted at true locations of the features they represent. Some attempts have been made to work around this problem on maps of urban areas such as generalizing some features, omitting features, or adding red overlays.

Streets depicted on the topographic map do not necessarily represent the actual street location or geometry. A neighborhood in Morgantown was chosen to illustrate this problem. This neighborhood was developed in the early 1900s and was designed to fit existing, moderately steep topography and to limit the number of through streets. Many streets have curves. Figure 4 shows that the DLG data derived from separates of the topographic maps do not agree with the imagery in several areas.

CONCLUSIONS

Geologic mapping is possible in much of north-central West Virginia using existing CBMP datasets; however, traditional fieldwork is critical in areas where little or no pre-existing data is available and/or new exposures are being created as development occurs.

Urban geologic mapping requires a degree of accuracy that is not generally regarded as necessary in rural geologic mapping. Using georeferenced digital orthophotography for basemaps is preferable to using topographic maps. Urban mapping requires the use of best available aerial photos and software to allow three-dimensional data development and editing.

REFERENCES

Fonner, R.F., Reynolds, J.H., Cole, G.A., Jake, T.R., Fedorko III, N., and Connell, D.A., 1981, Geologic cross section and apparent structure along the West Virginia portion of U.S. 48: West Virginia Geological and Economic Survey, 65 p.

The North American Commission on Stratigraphic Nomenclature, 1983, North American stratigraphic code: American Association of Petroleum Geologists Bulletin, v. 67, no. 5 (May, 1983), p. 841-875.

West Virginia Geological Survey, 1963, The West Virginia Geological and Economic Survey: Its accomplishments and outlook: West Virginia Geological and Economic Survey, 207 p.

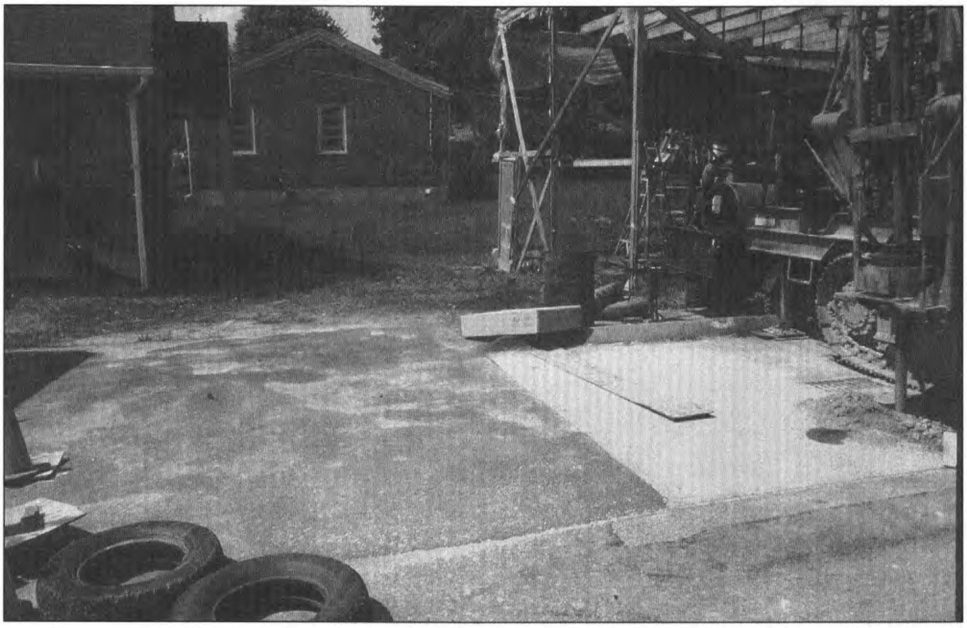


Figure 3. Subsidence damage and abatement effort at MacAurthur, West Virginia. Notice damage to foundation of brick building at left of photo.

Figure 4. Part of the South Park neighborhood, Morgantown, West Virginia. Georeferenced image produced by Positive Systems Corporation in 1997 overlaid with the street centerlines from the USGS DLG of the Morgantown South 7.5-minute topographic quadrangle. Both data sets have identical projections. Black arrows indicate some significant errors in the location of streets in the DLG.

Alaska Digital Mapping Project: From Field to Publication

By Christopher P. Garrity

U.S. Geological Survey
National Center, MS 956
12201 Sunrise Valley Drive
Reston, VA 20192
Telephone: (703) 648-6426
Fax: (703) 648-6419
e-mail: cgarrity@usgs.gov

INTRODUCTION

The natural resources of northern Alaska continue to be a focus of national attention. As a result, the need for detailed geologic maps is greater than ever, not only as a basis for petroleum and mineral exploration, but also for land-use planning and mitigation of environmental impacts. The U.S. Geological Survey (USGS), in cooperation with the Alaska Division of Oil and Gas, has embarked on a project to map the northern front and foothills of the Brooks Range of northern Alaska. The goal of the project is to publish an integrated set of digital geologic maps with consistent stratigraphic nomenclature and a standardized cartographic style. Initial field work in the Alaska North Slope, including the northern front and foothills of the Brooks Range, was accomplished between 1944 and 1953 by the USGS. Maps stemming from that work were published between 1957 and 1966 as USGS Professional Papers 303 A-H. Since then, numerous geologic maps of individual quadrangles at various scales have been published by the USGS and by the Alaska Division of Geological and Geophysical Surveys. These maps provide a broad framework of the stratigraphy and general outcrop distribution of northern Alaska. The current project builds from these maps, incorporating new field observations, modern stratigraphic terminology, and contemporary digital mapping techniques (for example, Mull, 2003).

This paper discusses techniques employed in digitally capturing the geologic maps of northern Alaska and the post-capture procedures used in preparing publications. A variety of Geographical Information Systems (GIS) and graphical software applications, including ArcGIS, Adobe Illustrator, Adobe Photoshop, and Avenza MAPublisher were used in the cartographic production process.

IMAGE CAPTURE AND PREPARATION

For this project, original geologic linework and subsequent map revisions were drafted on stable-base mylar, which substantially reduced the swelling, shrinking, and folding errors commonly associated with scanning paper maps. Before stable-base mylars were scanned, the scanner was serviced to assure proper camera alignment and roller control. Mylars were then scanned at a resolution of 400 dots per inch (dpi) and saved as an 8-bit grayscale Tagged Image File Format (TIFF) image. The improved clarity gained by scanning at a resolution greater than 400 dpi did not merit the resulting increase in file size.

Post-scan image filtering was done in Adobe Photoshop. The first filter applied to grayscale TIFF images was "unsharpen mask." This filter identified contrast edges in the image and increased the overall contrast, which performed exceptionally well in defining geologic contact lines. The "amount" was set to 200%, with a radius of about 1.5 pixels and a threshold level of 0%. Applying the filter lessened the amount of image "wash-out" when the grayscale image was later converted to monochrome. The next step, easily the most time-consuming part of the digital capturing process, was to remove the non-essential elements in the image (geologic unit labels, strike and dip symbols, leaders, etc.). Eraser tools in Adobe Photoshop and raster-cleanup tools in ArcScan were used for this. Finally, the image was converted from 8-bit grayscale to 1-bit monochrome, using the 50% threshold option (fig.1). This option converted pixels with gray values above the middle gray level (value=128) to white and converted gray levels below the middle level to black. A thorough visual analysis of the image was performed to ensure that features were not washedout upon conversion. Conversion to monochrome was useful, as ArcScan requires images to be bi-level (two unique colors) prior to vectorization. An additional benefit associated with the monochrome conversion was a significant reduction in file size.

The final step before vectorization was to georeference the image using the georeference tools in ArcMap. The image was georeferenced by adding Universal Transverse Mercator(UTM) coordinates to each latitude and longitude tick-mark on the source map. After an ac-

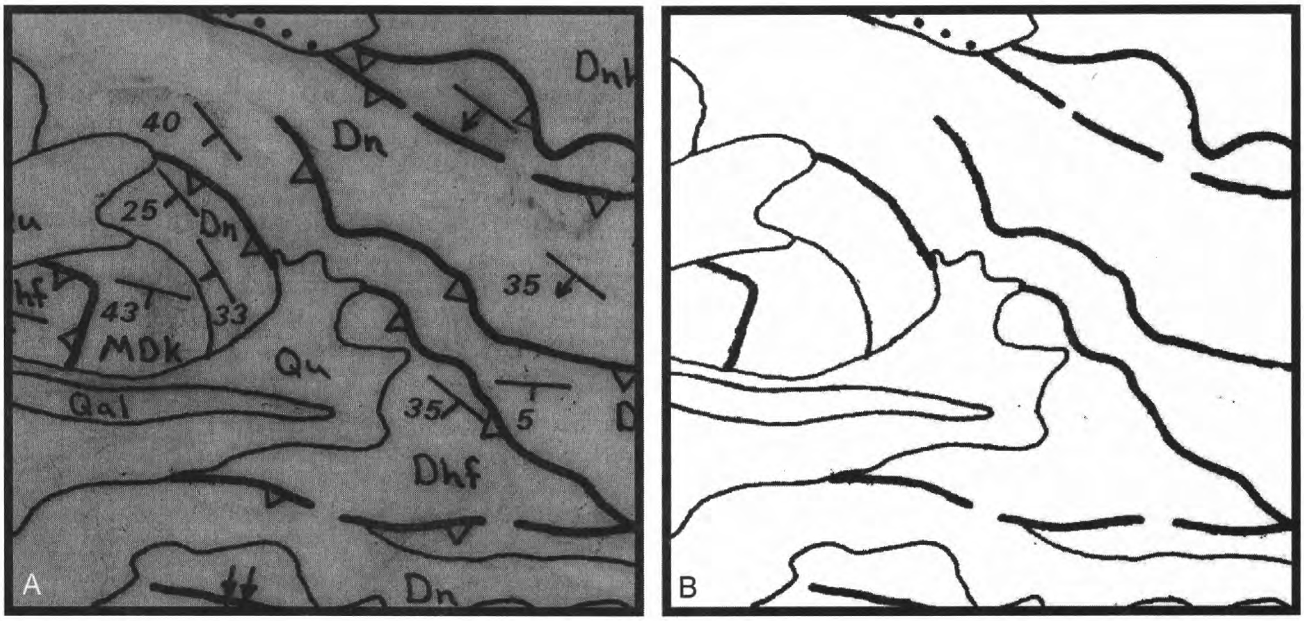


Figure 1. Preparing images for autovectorization. (A) Original grayscale image scanned at 400 dpi. (B) Monochrome image after unwanted raster elements were removed.

ceptable Root Mean Square error value was attained, the image was rectified and added to the data frame.

STABLE-BASE VECTORIZATION

TIFF images were autovectorized using the ArcScan extension in ArcGIS 8.3. Almost all raster-to-vector conversions were accomplished through batch (automated) vectorization. This method was less time-consuming than interactive vectorization, although sections of numerous maps were nevertheless vectorized interactively. Problem areas encountered in the batch vectorization process were repaired with the suite of raster painting tools available in the raster painting toolbar (fig. 2).

The geometry of the resulting vector data was controlled in the vectorization settings menu. To maintain the integrity of the source linework, special attention was paid when adjusting settings. All maps were vectorized using the centerline vectorization method. Vectorization settings used for digital capture of source scans included the following:

- Intersection solution was set to median; median is designed for nonrectilinear intersections (ideal for geologic maps).
- Maximum line-width setting varied by scan, but was used to extract vectors representing the same feature in the source map (for example, faults or geologic contacts).
- Compression tolerance was set between 0.005 and 0.050, depending on the complexity of the source map. This produced smooth vectors and represented the source linework accurately. When figures with angular intersections were vectorized, compression tolerance was increased significantly to generate straight-line vectors with a limited number of vertices.
- Smoothing weight was set between 1 and 3. When levels were set higher than 3, misinterpretation of the source data commonly occurred.
- Gap closure tolerance was used when inferred or concealed line symbols were denoted in the source map. When the setting was activated, vectors would sometimes connect to random speckles in the image, generating unwanted line features. For this reason, all source images were despeckled before vectorization.

By using the maximum line width setting and raster-cell selection tools, vector data representing the same feature were extracted and saved into separate data layers. This proved to be helpful when vector data were attributed. Separate data layers were saved in ArcInfo coverage format and integrated into the digital database. Finally, ArcInfo export files were generated for cartographic production in Adobe Illustrator via MAPublisher.

FILM POSITIVE BASES

For each quadrangle, topography, drainage, and open-water film positives for base-map compilation were ordered through the USGS Eastern Publications group. Unfortunately, open-water separates were not available for every quadrangle, and in some instances had to be generated by autovectorizing drainage positives. In these

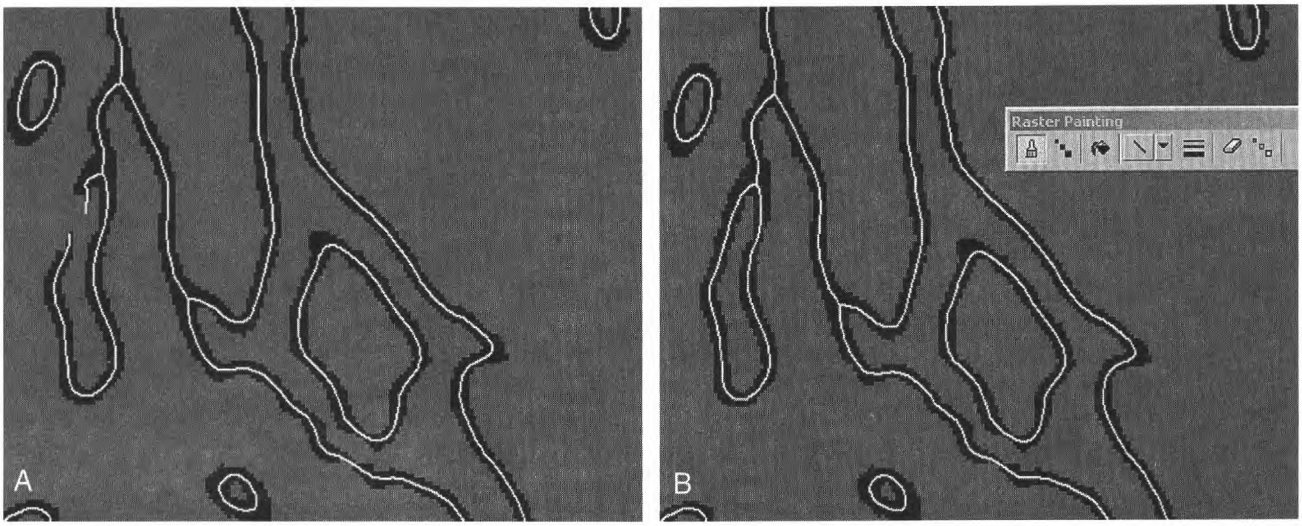


Figure 2. Raster cleanup in ArcScan. (A) A section of the raster image (black pixels) failed to be detected by the scanner. (B) The section was repaired with the raster painting brush, and the vector preview (white pixels) was generated interactively.

cases, the same scanning procedures discussed above were applied. After polygon topology was built, open-water polygons were extracted and saved as a new coverage. For all other separates, scanner threshold settings were adjusted and film positives were scanned directly to monochrome. A monochrome output was chosen because all background pixels (white fill areas) appear transparent when opened in Adobe Illustrator. Base map layers could then be placed over other graphic layers without concealing them. Positives were scanned at a resolution of 600 dpi to avoid legibility problems in areas where linework was very fine. Image collar information was trimmed and speckles were removed before being georeferenced in ArcMap. Georeferenced images were positioned accurately relative to underlying vector geology using the MAPublisher register image filter. Base images were placed as links in Illustrator so that future image editing could be accomplished "on the fly." Color choices for base-map separates were as follows:

- A 50 – 65% gray level was applied to the topography/culture separate. Gray levels were adjusted on the basis of the complexity of the raster image. When levels above 65% were applied, the raster image began to obscure underlying vector data and was generally too overbearing. When gray levels below 50% were applied, the raster image suffered from diminished legibility in areas where underlying colors were light.
- A 100% cyan level was applied to the drainage separate. This distinguished the drainage separate from the open-water separate, which was assigned a cyan level near 25%. In areas where underlying layers had a high cyan content, image color levels were adjusted to avoid legibility problems.

CARTOGRAPHIC DESIGN

ArcInfo export files were imported into Adobe Illustrator via MAPublisher for cartographic production. In some cases, arcs with too many vertices created diced polygons upon import (fig. 3). Although setting the grain tolerance in MAPublisher can resolve this dilemma, in areas where problems remained, arcs were generalized in ArcInfo. After coverages were imported, all unique features were released into separate layers for greater ease in editing and review. By importing via MAPublisher, attributes were retained, allowing unique features to be selected with the "select by attribute" feature. This was quite helpful when stroking line features, using USGS digital cartographic standards (U.S. Geological Survey, 2000). Federal Geographic Data Committee (FGDC) compliant cartographic specifications in PostScript format were easily implemented in Adobe Illustrator with the eyedropper tool. Line features requiring uniform symbols along their axes (thrust fault teeth, hachure marks, etc.) were built as pattern brushes in Adobe Illustrator to ensure accurate symbol placement without the need to place symbols individually. Point features requiring rotation, such as strike and dip, were rotated in ArcGIS via a rotation field and exported in PostScript format. Collar text information, including correlation of map units, description of map units, discussion, and references, was pasted into text boxes for ease in standardization. Finally, drafts were sent to the USGS Eastern Publications Group for technical review and editing.

Completed digital quadrangles have been converted to Printable Document Format (PDF) files for future online release. Ultimately, through the World Wide Web, users will be able to download the PDF files, ArcInfo coverages, or order hard copies through the map-on-de-

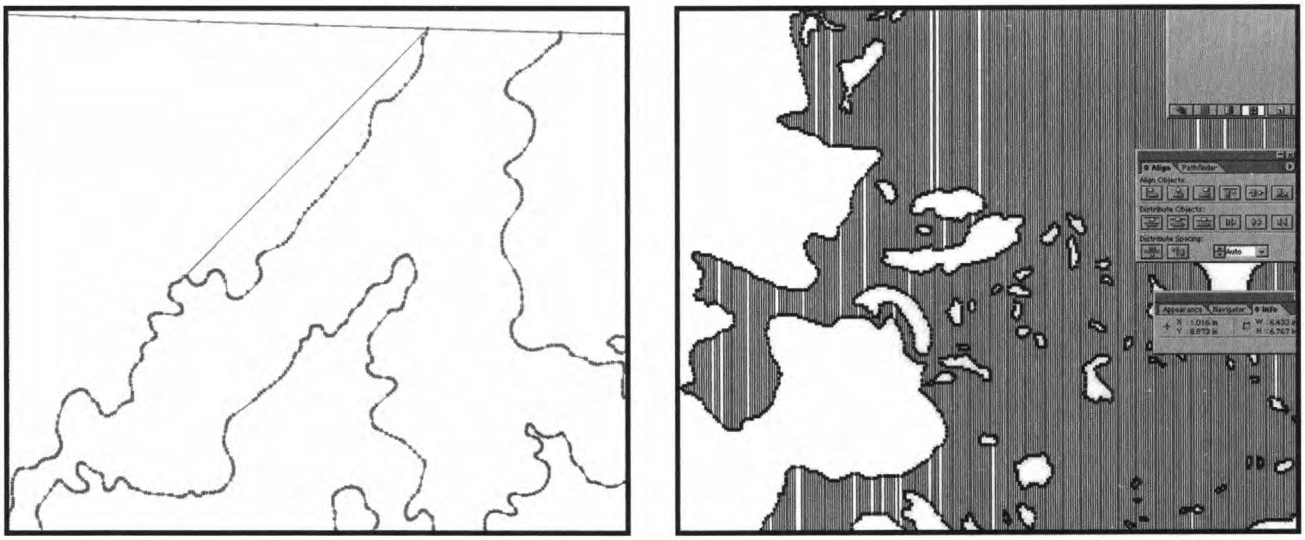


Figure 3. Two examples of diced polygons in MAPublisher. This occurs when arcs have too many vertices, and is easily resolved by generalizing problem arcs in ArcInfo.

mand process. Plotting-on-demand eliminates the need to make expensive color-separated negatives, but generally does not meet the same quality and accuracy standards as lithographically printed products.

REFERENCES

Mull, C.G, Houseknecht, D.W., and Bird, K.J., 2003, Revised Cretaceous and Tertiary stratigraphic nomenclature in the Colville Basin, northern Alaska: U.S. Geological Survey Professional Paper 1673, <http://pubs.usgs.gov/pp/p1673/>.

Mull, C.G, Houseknecht, D.W., and Pessel, G. H., in press-2003, Geologic map of the Umiat quadrangle, Alaska: U.S. Geological Survey MF map, scale 1:250,000.

U.S. Geological Survey, 2000, Public Review Draft – Digital cartographic standards for geologic map information: U.S. Geological Survey Open-File Report 99-430, p. 1-18, <http:// geopubs.wr .usgs.gov/open-file/of99-430/>.

A Map of Lightning Strike Density for Southeastern Pennsylvania, and Correlation with Terrain Elevation

By Alex J. DeCaria and Michael J. Babij

Millersville University
Department of Earth Science
P.O. Box 1002
Millersville, PA 17551-0302
Telephone: (717) 871-4739
Fax: (717) 871-4725
e-mail: alex.decaria@millersville.edu

INTRODUCTION

The National Lightning Detection Network™ (NLDN) consists of 106 sensors in a nationwide array (Orville and others, 2002) that detect and locate cloud-to-ground (CG) lightning strikes. Complete national coverage (with varying detection efficiency) began in 1989. The data from the NLDN are proprietary, and are controlled and distributed by Väisälä, Inc., Inc. Though extensive data sets must be purchased from Väisälä, Inc., the company makes limited data sets available for research at educational institutions at no charge. We were able to obtain 7 years worth of data (1995-2001) for southeastern Pennsylvania with the intent of comparing annual average lightning strike density with terrain elevation.

CREATING THE LIGHTNING DENSITY MAP

The lightning data were obtained in ASCII format, and included the position, in latitude and longitude referenced to the WGS 84 datum; the time of the strike, in Universal Time Coordinate, or UTC; and the signal, either negative or positive, depending on whether the strike lowers negative charge or positive charge to the ground, respectively. A Fortran 90 program was written to read the input file and write separate ASCII files for each year. The ASCII files were then opened using Microsoft Access database software to create database files readable by ArcGIS software. The Microsoft Access database files were read into ArcGIS, and from them, shape files were created for each year. Figure 1 shows the locations of all 80,487 lightning strikes recorded in Southeastern Pennsylvania during 2001 (only 1 year is shown because data for multiple years is too dense to show here). The county outlines were prepared by the Pennsylvania Department of Environmental Protection from data obtained from the Pennsylvania Spatial Data Access (PASDA) website, http://www.pasda.psu.edu.

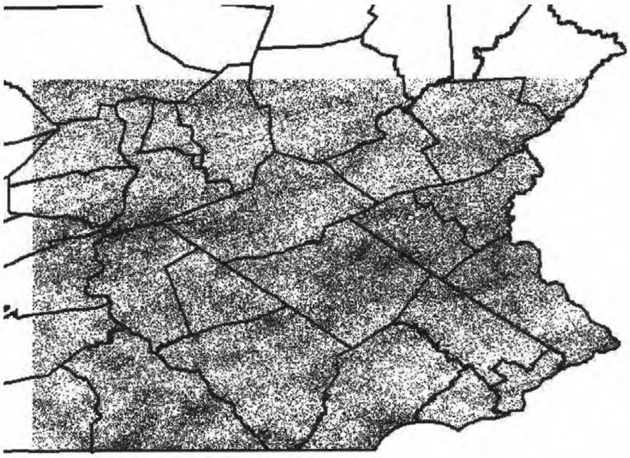


Figure 1. Location of all 80,487 lightning strikes recorded in southeastern Pennsylvania during 2001.

Once the lightning strike positions were put into ArcGIS, the Spatial Analyst extension was used to calculate the strike density. The desired output units were number of strikes per square kilometer. Prior to calculating the density, the coordinates and projection were converted to an Albers equal area projection with coordinates in meters. A cell size and search radius of 1000 meters were used for the density calculations (this cell size was deemed suitable based on a conversation with Ken Cummins of Väisälä, Inc.). The strike densities were then corrected for the estimated detection efficiency of the NLDN for Southeastern Pennsylvania (provided by Väisälä, Inc.). The detection efficiency in 1995 was 61%, and increased to 80% in 1996 and 85% in 1997. From 1998 onward the estimated detection efficiency was 100%. After the strike densities for individual years were computed, they were averaged using the raster calculator feature of the Spatial Analyst extension to ArcGIS. Figure 2 shows the computed lightning strike density per year per square kilometer for 1995-2001.

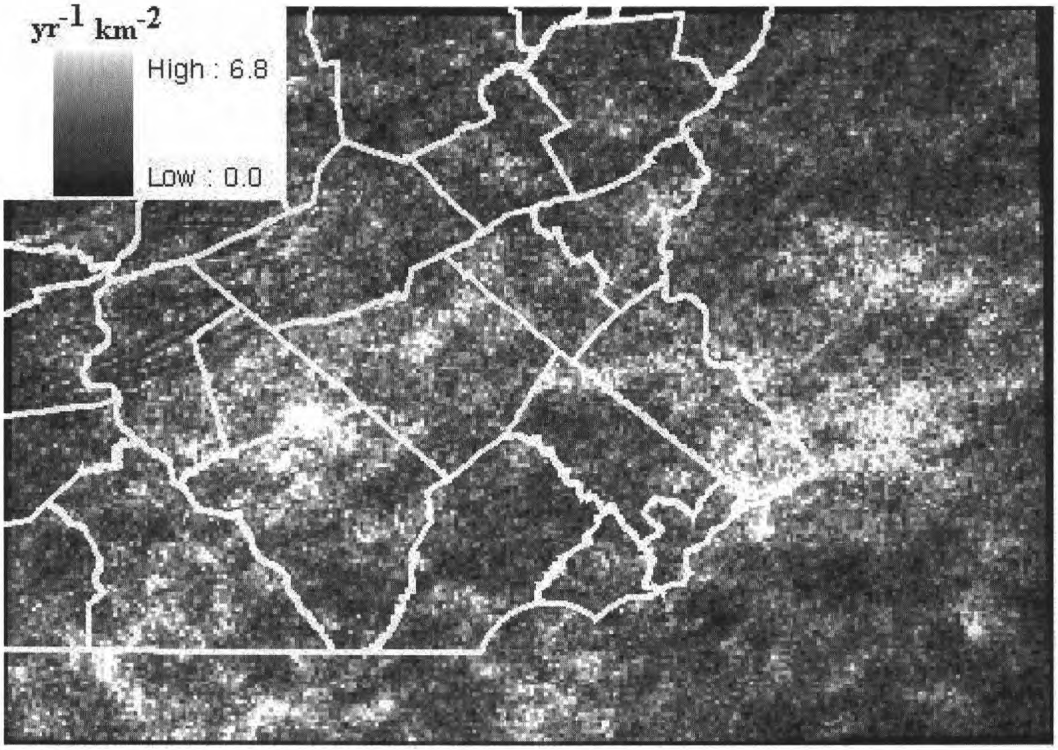


Figure 2. Computed lightning strike density in flashes per square kilometers per year (yr^{-1} km^{-2}) for southeastern Pennsylvania from 1995-2001.

CREATING A JOINED DATA SET OF LIGHTING STRIKE DENSITY AND TERRAIN ELEVATION

For comparison of strike density with terrain elevation, elevation data at a horizontal resolution matching that of the strike density data (1 km) were needed. The desired elevation data source was the U.S. Geological Survey (USGS) National Elevation Dataset (NED) provided by the Pennsylvania Topographic and Geologic Survey. Initially, this data set was resampled to a lower resolution by first converting the raster elevations into vector (point) format, and then converting back into a lower-resolution raster. This procedure overly taxed the memory of the personal computer used for the analysis, and caused the software to lock up. Therefore, a lower-resolution (~100 m) digital terrain model prepared by the Pennsylvania Department of Environmental Protection was downloaded from the Pennsylvania Spatial Data Access website and resampled to 1-km resolution (fig. 3). This difficulty was a direct result of our inexperience with the ArcGIS software. We have since learned a simpler and more efficient method of re-sampling the data to a lower resolution while remaining in raster format, and this has worked successfully on the USGS NED.

The location of the data points in the vector strike density data set and the vector elevation data sets were not coincident. In order to compare elevation with strike density the two vector data sets were first clipped to a common domain, and then joined using an ArcGIS function to interpolate the two vector data sets to a common set of points.

ANALYZING THE CORRELATION BETWEEN LIGHTING STRIKE DENSITY AND TERRAIN ELEVATION

Using the joined vector data set of lighting strike density and terrain elevation, the correlation between strike density and elevation was analyzed using scatter plots and standard statistical methods. The attribute table from the joined density-elevation data set was exported as an ASCII file and read into Microsoft Excel, from which the statistical analysis was performed.

Figure 4 shows a scatter plot of strike density versus terrain elevation for the entire domain. There is a weak, yet statistically significant, negative correlation between strike density and terrain elevation. The lack of a positive correlation between strike density and elevation implies that terrain elevation is not the dominant factor for thunderstorm formation in this region. There is a significant maximum in lightning strike density in Bucks County, which lies along the Delaware River and is relatively low and flat. Mechanisms such as the convergence associated with the bay breeze coming up the Delaware River Valley, or uplift due to frontal passages,

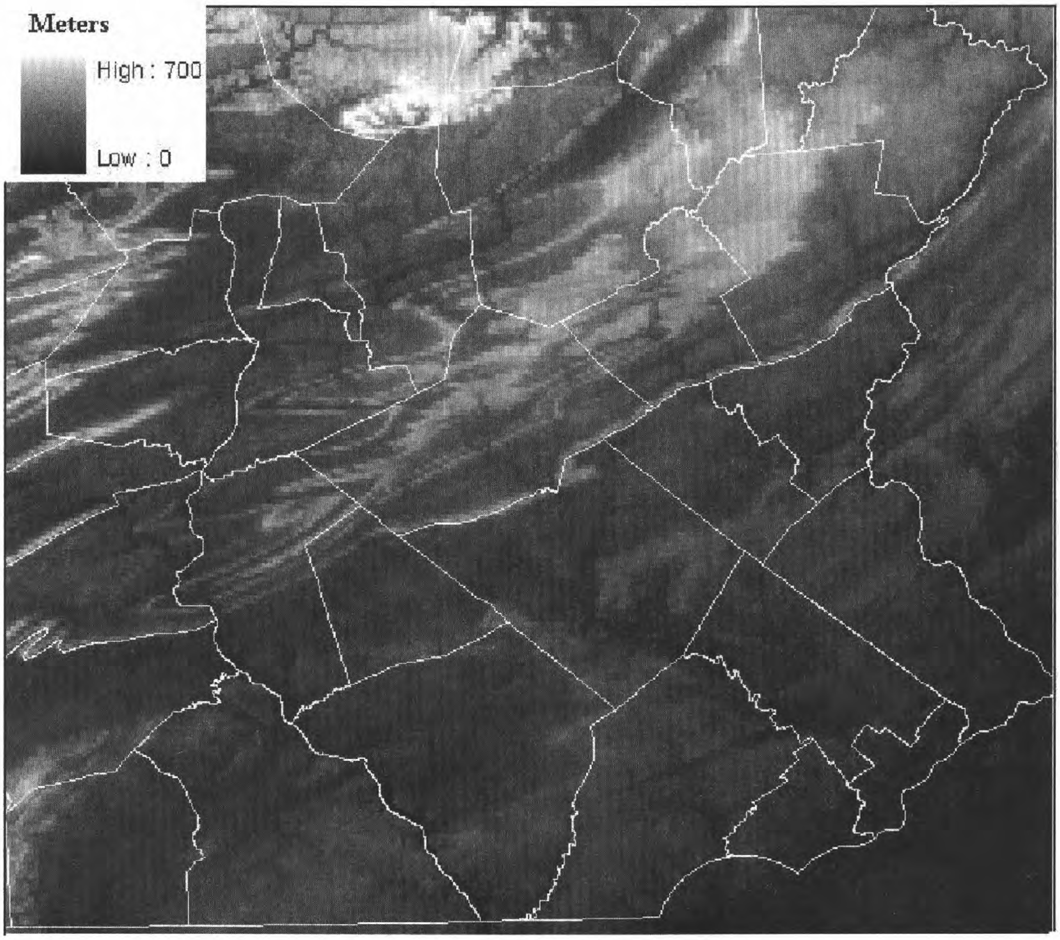


Figure 3. Low horizontal resolution (1 km) terrain for southeastern Pennsylvania. Units are meters.

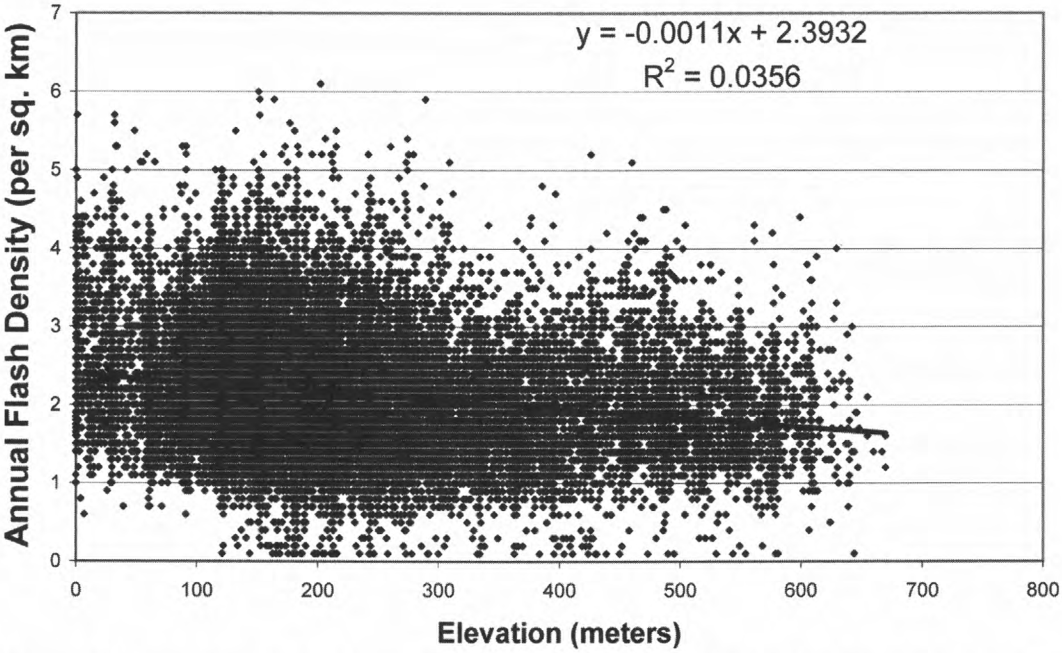


Figure 4. Scatter plot of strike density (yr^{-1} km^{-2}) versus terrain elevation (m) for southeastern Pennsylvania. The negative correlation, though weak, is statistically significant.

appear to be more significant for triggering thunderstorms in this region. This is consistent with Orville and Huffines (2001), who found a minimum in lightning strikes over the high terrain of the Appalachian Mountains. Scatter plots were also created for several individual counties, and in no case was a positive correlation found between strike density and terrain elevation.

SUMMARY

ArcGIS was used to manipulate lightning strike position data and terrain elevation data so that the correlation between lightning strike density and terrain elevation for southeastern Pennsylvania could be explored. A statistically significant, weak negative correlation between lightning strike density and terrain elevation was found in southeastern Pennsylvania. This indicates that orography is at best a secondary factor for the formation of thunderstorms in this region, since we would expect to see a positive correlation if the thunderstorms were orographically forced.

ACKNOWLEDGMENTS

We would like to acknowledge Väisälä, Inc. for providing the lightning strike data free of charge for this research project. We are also thankful to Tom Whitfield of the Pennsylvania Topographic and Geologic Survey for providing elevation data and also advice and guidance regarding use of ArcGIS. This research was funded under a Millersville University Faculty Research Grant.

REFERENCES

Orville, R.E. and Huffines, G. R., 2001, Cloud-to-ground lightning in the United States: NLDN results in the first decade, 1989-98: Monthly Weather Review, v. 129, p. 1179-1193.

Orville, R.E, Huffines, G.R., Burrows, W.R., Holle, R.L., and. Cummins, K.L, 2002, The North American lightning detection network (NALDN) – First results: 1998-2000: Monthly Weather Review, v. 130, p. 2098-2109.

Negligence and Professional Malpractice Related to GIS Datasets

By Ian J. Duncan

Texas Bureau of Economic Geology*
University Station X
Austin, Texas, 78713
Telephone: 512-471-1534
Fax: 512-471-0140
e-mail: ian.duncan@beg.utexas.edu
*Formerly at Virginia Department of Mines Minerals and Energy

INTRODUCTION

Chrisman (1991) has suggested that "error (in spatial data) is inescapable, it should be recognized as a fundamental dimension of data." Digital geologic data sets are not an exception to this truism. Typical errors include incorrect data, missing data, incorrect georeferencing, and incorrect documentation of the data. Although these types of errors can always occur, well-established methods are available to characterize them. Informing users of the data's reliability and the nature of errors in the dataset can contribute greatly to effective use of the data. Geologists and GIS professionals should develop and implement a comprehensive approach to addressing these issues.

Liability can arise when a person or company alleges harm (or an adverse outcome) resulting from a bad decision made on the basis of erroneous data. Liability and negligence can be related to contract law (Schultz, 1999; West Publishing Company, 1998), to common law (the law of torts, Phillips, 1996), or to legislative statutes (West Publishing Company, 1998; Creenan, 2003). Karnow (1996), a trial lawyer, has succinctly summarized the liability associated with errors: "When mistakes are made, one simply traces back the vector of causation to the negligent human agency that caused the error. Then in theory one sues that (human) for damages and is made whole. The sins of omission and commission are just as subject to condemnation as negligence, recklessness, malfeasance, or other human culpability.

This paper examines the nature of professional negligence and malpractice, and the liability exposure of individual geologists and GIS professionals related to their production of digital spatial data. Negligence is the failure to exercise reasonable care (a standard of care normally expected of someone engaged in that type of work), when harm results from this failure. Kauffman (1994) has reviewed the exposure of public officials in the U.S. to negligence charges and provides a useful overview of the situation. Onsrud (1999) has specifically addressed the role of negligence in assessing liability associated with GIS data sets, noting that: "Negligence is conduct that breaches a duty of care for the protection of others against unreasonable risk of harm. Each person owes a duty to act as a reasonable person would under same or similar circumstances. Thus, negligence does not involve intent to cause harm but only a failure to meet a sufficient degree of care." According to Creenan (2003) "the duty owed by the professional can be determined by agreement, law, or by the standard of care applicable to the profession… based on specialized education, knowledge and skill."

Onsrud (1999) defines malpractice as "when an individual does not perform to the expected standard of his profession." In addition, Onsrud asserts that "The greater the risk of harm involved the greater the care that a reasonable person can be expected to take. Under negligence, the duty to act reasonably to avoid foreseeable risks of physical injury extends to any person." Onsrud concludes that "in the U.S., it is probably safe to conclude that Government GIS offices assume at least some liability exposure in collecting and disseminating land-related data." Donohoe (2000) has reached a similar conclusion regarding liability exposure of professionals working for the government, noting that "the basic rule (in Australia) is that the tort liability of public authorities is governed by the same principles that apply to private individuals." Onsrud (1999) has pointed out the decline of sovereign immunity in the U.S. Sovereign immunity is a doctrine precluding law suits against the sovereign [government] without its consent. However, a growing body of case law suggests that individuals working for the government are becoming increasingly subject to tort actions (Doyle, and Redwood, 1999).

Perhaps surprisingly, the possibility of negligence related to digital geologic data may be greater than that related to the same data in paper form. Tarter (1992) has noted that "(the) myth of machine infallibility seems to create a demand for a higher standard of quality for machine readable data than for traditionally distributed information." Similarly Peritz (1986) has suggested that "the presumption of trustworthiness (of digital data) simply carries too much weight in our recently computerized society." As a consequence, courts may entertain claims of liability, negligence, or malpractice more readily if data is in digital form and indeed, if it is supplied via a GIS. Anderson and Stewart (1995) have observed "with a growing clientele that increasingly expects data from the computer to be 'right', the potential for exposure to a (law) suit by disgruntled customers is growing."

Standard of Care: The Concept of a Reasonable Person

As summarized by West Publishing Company (1998), under U.S. law "a person has acted negligently if he or she has departed from the conduct expected of a reasonably prudent person acting under similar circumstances. The hypothetical reasonable person provides an objective by which the conduct of others is judged. In law, the reasonable person is not an average person or a typical person but a composite of the community's judgment as to how the typical community member should behave in situations that might pose a threat of harm to the public." The concept of a "reasonable person" (or a "reasonable man") can be traced back to the 1856 case of *Blyth v. Birmingham Water Works* in which the judge concluded:

> "Negligence is the omission to do something which a reasonable man, guided upon those considerations which ordinarily regulate the conduct of human affairs, would do, or doing something which a prudent and reasonable man would not do. The defendants might have been liable for negligence, if, unintentionally, they omitted to do that which a reasonable person would have done, or did that which a person taking reasonable precautions would not have done."

The standard of a "reasonable person" has never been clearly defined and this is perhaps the great strength of the concept. As pointed out by the judge in *Carlson v. Chochinov* (1947) (quoted in Duhaime, 1996) "the ideal of that person exists only in the minds of men, and exists in different forms in the minds of different men. The standard is therefore far from fixed or stable. But it is the best all-round guide that the law can devise."

The standards of care expected of professionals are higher than those imposed on an average citizen. A professional is expected to demonstrate "due diligence" and a minimum level of expertise and performance determined by norms for a particular profession. Duhaime (1996) states "Compliance with customs, such as professional customs, generally will exonerate a defendant as it provides excellent proof of what is "reasonable" conduct. But, beware: judges are fully enabled to find a custom as lacking in providing reasonable care and, therefore, so would the conduct that may have been guided by that custom." A new South Australian negligence law reform proposal (Turner and Groom, 2003) suggests the following criteria: "A person who provides a professional service incurs no liability in negligence if... the provider acted in a manner that was widely accepted in Australia by members of the same profession as competent professional practice." Creenan (2003) has pointed out that even in a state with strong professional licensing statutes (Pennsylvania) the professional standard of care is gauged by expert witnesses and not by statute. There is only limited case law relevant to malpractice or negligence related to GIS data files or geologic maps. This lack of judgments may be misleading as most such cases are likely to be settled out of court.

STRATEGIES TO LIMIT LIABILITY ASSOCIATED WITH DIGITAL DATA

Anecdotal and some survey data reveal that attaching disclaimer statements to digital data sets is the main strategy used by geologists and GIS professionals to address liability issues related to the data. Disclaimer statements are certainly proliferating on GIS and Geospatial Data Clearing House Web sites. The following disclaimer statements associated with digital geologic data sets were found on the Internet:

- *You may use the digital maps made available from this agency if you assume complete legal and ethical responsibility for the problems resulting from their use.*
- *NO WARRANTY: The (agency) provide these GIS data "as is". The (agency) makes no guarantee or warranty concerning the accuracy of the information. The (agency) makes no warranties, expressed or implied, as to any matter whatsoever, including but not limited to the condition of the product or its fitness for any particular purpose. The burden for determining fitness of use lies entirely with the user.*
- *LIABILITY: The entire risk as to the usage of this data is assumed by the end user.*
- *The (agency) makes no representations or warranty expressed or implied with respect to its accuracy, completeness, or usefulness for any particular purpose or scale. The agency assumes no liability for damage resulting from the use of any information, method or process disclosed in this data set, map or text, and urges independent site-specific verification of the information contained herein.*

Only the last of these disclaimer statements gives the end user any hint as to how they might evaluate their risk in using the data, by suggesting an independent site specific investigation.

Dansby (1992) has used case law to examine the liability related to making GIS data available to the public. Dansby's arguments suggest that disclaimers in general are likely to be ineffective as protection against negligence or malpractice law suites. Fogl (1998) and Reid et al (1996) have reached similar conclusions.

Rather than disclaiming responsibility, statements directed at customers are best used to clarify their expectations of the data. Dansby (1992) noted that explanations of data accuracy and its limitations on usability can be effective (though such explanations are not disclaimers in a legal sense). Such warnings modify the reliance and reasonable expectations of those receiving the data. A geologist or GIS professional cannot be held responsible for the inevitable errors that occur in any real world database. However, a professional does have a responsibility to prevent systematic patterns of error, data poorly characterized as to error or other aspects of quality, or data that somehow gives the user a false or misleading impression of its quality.

Development of a Customer Friendly "Disclaimer"

Autenucci and others (1991) have concluded that to avoid legal liability for GIS data "the objective of a map disclaimer (should be) to alert users that errors are possible and that they need to consider the degree of reliance appropriate for their purpose". Traditionally the metadata associated with digital data has been the mechanism by which users were made aware of the possible errors in data sets. Metadata can play a key role in management of liability. For example, the Ventura County California Planning Division's Metadata Policy states: "from the (data) producer's perspective metadata is a means of declaring data limitations and serves as a form of liability insurance."

To follow up on Dansby's (1992) suggestion that the explanation of the limitations of the data set is more useful than a disclaimer, the following statement was developed by the author to integrate aspects of metadata information with a customer friendly statement:

This data set was created for a particular regulatory or scientific purpose and cannot be expected to have the accuracy, resolution, completeness, or other characteristics appropriate for all the applications that potential users of the data may contemplate.

A comprehensive effort has been made to characterize the accuracy, resolution, and completeness of this data set. The statistical measurement of error and the estimates of resolution and completeness are detailed in the metadata. Also, included in the metadata is a review of the methodologies used to measure and/or estimate these parameters.

This data is supplied in an attempt to support economic development, assure public safety, and preserve the environment. Before using this data, you should read the following summary of how to use the accuracy, resolution, and completeness information (contained in the metadata) to assess the "fitness of use" of this data set for your application. You should assess decide whether the documented reliability of this data set meets the needs of your particular application. You may contact us to help you in your evaluation. In particular, please consider:

Accuracy—The accuracy of each entity in this data set has been estimated by taking a random sample and remeasuring the parameters using a more accurate method such as GPS for location. The accuracy is computed based on the National Spatial Data Standard and is expressed at the 95 percent confidence level. (You should consider how the accuracy of this data compares with the requirements of your application)

Resolution—The resolution of the data is defined as the smallest entity recorded in the data set. (You should consider the resolution needs of your application in the context of the resolution of this data).

Completeness—The completeness of this data set has three components:

(i) The percent completeness of the data set with respect to area covered, or to the number of entities captured versus entities in existence (as determined by statistical sampling).
(ii) The completeness of categorization of the data. For example, a land-use map that consisted of five categories and lumped 40 percent of land use types into an "other land use" category, would be incomplete.
(iii) The completeness of verification of the data. This measure of completeness addresses the extent to which the data has been checked by some specified process.

We encourage you to contact us to collaborate with you in making a "fitness of use" evaluation. By adding the results of this analysis to our metadata, we hope to enable other potential users of this data to understand its strengths and limitations.

To be effective, such a statement should direct end users toward information that characterizes the data in both a quantitative and qualitative fashion. In addition, geologists should supply concrete examples and/or methodologies that assist the end user in assessing the fitness of the data for a particular use.

DISCUSSION AND CONCLUSIONS

Unfortunately, much geologic data was collected for a limited purpose and do not have the accuracy or completeness that many end users would prefer or need. For example, end-users conducting field work based on high accuracy GPS devices may be using geologic data collected on a topographic map base with far less accuracy. To address such problems, efforts should be directed towards adequate characterization of the accuracy, the resolution, and the completeness of the data sets. By characterizing the quality of the data in such a way that the end user can understand its limitations, a geologist or GIS professional can reduce personal liability and increase the utility of the data. Unfortunately most data producers have chosen to rely on disclaimers rather than attempting to use metadata to limit liability. For example in Mason and Masters' (1999) survey of spatial data producers in Australia, 64 percent supplied disclaimers and only 34 percent supplied metadata.

Onsrud (1999) has suggested that judges recognize that mistakes and blunders are inevitable. Error-free data, ideal for all applications, is not the court's expectation. Rather, courts are likely to ask questions in the form:

- Are the data collected and assembled in a competent manner, reflecting best practices of the profession?
- Is a good faith effort made to characterize the data quality in terms of accuracy and completeness?
- Are data quality information presented in such a way that a reasonable person could be expected to gauge the appropriateness of the data for a particular purpose?
- Are nationally accepted data formats for metadata standards followed?
- Are the standards implemented in a reasonable and competent manner?

Uncertainty in the information used to make decisions implies a risk associated with such decisions. This is the nexus between uncertainty analysis in GIS and risk management. To reduce the risk, the key aim has to be to increase the quality of the decisions based on the data. To achieve this, it may be more important to improve the characterization of the data than to improve the spatial accuracy of the data. The risk of inappropriate usage increases as digital geologic data becomes common in decision-making.

For a geologist or GIS professional, whose job involves the creation and distribution of digital data sets, addressing liability issues should be a major component of a personal risk management plan. Strategies to minimize the possibility of negligence actions should include:

- Creating comprehensive metadata, for all datasets that you produce, that meet or exceed national standards.
- Facilitating the effective use of metadata by customers through education and training.
- Educating yourself in appropriate aspects of quality management, quality measurement, and liability.
- Keeping accurate records on how data is produced and characterized. Following Kauffman (1994), ask the question "if taken to court, how can I prove that I did what I said I did?"
- Consider acquiring personal malpractice insurance (Epstein et al, 1998) if you have a position with clear responsibility over datasets.

In conclusion, professional scientists involved in producing digital spatial data have a professional exposure to negligence and malpractice suits. Recent reviews of the case law have suggested that professionals can be held accountable for the accuracy and reliability of digital spatial data and can be found negligent for failure to take sufficient care. Authors of digital data are responsible for adequately characterizing the data's reliability, accuracy and completeness. The standard of care expected is that of "due diligence" and the minimum level of professional care determined by norms for a particular profession.

REFERENCES

Anderson, J. R., and Stewart, A.R., 1995, Local government liability for erroneous data, *in* Proceedings of the Conference on Law and Information Policy for Spatial Databases NCGIA, p. 267-279

Autenucci, J., Brown, K., Croswell P. and Kevany, J., 1991, Geographic Information Systems: New York, Chapman Hall, _p 257 .

Chrisman, N., 1991, The error component in spatial data, *in* Longley, P., Goodchild, M. F., and Maguire D. eds., Geographical Information Systems, London, Longman, p. 165-174

Creenan, J. W., 2003, New Rules for Professional Liability Claims: www.waymanlaw.com/pdffiles/WIM_GWD.WIM_PO.WIM_LIB_191327.pdf

Dansby, H.B., 1992, Public Records and Government Liability: GIS Law, v.1, no. 2, p. 8-13

Donohoe, P., 2000, Trends in Liability of Public Authorities in Torts: Plaintiff, v. 40, August, p. 21-26

Doyle, J., and Redwood J., 1999, The Common Law Liability of Public Authorities: Tort Law Reviews, v. 7, no. 1, p. 30-51

Duhaime, L., 1996, Negligence - An Introduction: www.duhaime.org/Tort/ca-negl.htm

Epstein, E. F. et al, 1998, Liability Insurance and the use of Geographical Information: International Journal of Geographical Information Science, v. 12, no. 3, p. 203-214

Fogl, R., 1998, Disclaimers: Can they Limit Liability for Misleading and Deceptive Conduct: Charter, v. 69, no. 4, p. 93

Karnow, C., 1996, Liability for Distributed Artificial Intelligence: Berkeley Technology Law Journal, v. 11, Spring, available at www.law.berkeley.edu/journals/btlj/articles/vol11/Karnow/html/text.html

Kauffman J., 1994, Public Officials: Errors or Omissions: Professional Safety, v. 39, no. 9, p. 29-32

Mason, R., and Masters, E., 1999, National Spatial Data Survey: Azimuth, May, report available at http://129.94.250.112/survey/report

Onsrud, H. J., 1999, Liability in the use of GIS and Geographical Data Sets, *in* Longley, P., Goodchild, M. F., and Maguire D. eds., Geographical Information Systems, v. 2: London, Longman, p. 643-652

Peritz, R., 1986, Computer Data and Reliability: North West University Law Review, v. 80, p. 960

Phillips, J., 1996, Information Liability: The Possible Chilling Effect of Tort Claims Against Producers of Geographic Information Systems Data: Florida State University Law Review, v. 26, no. 3, p. 743

Reid, K., Clark, E., and Cho, G., 1996, Legal Risk Management for GIS: Journal of Law and Information Science, v. 7, Issue 2, available at http://www.jlis.law.utas.edu.au/v7i2risk_GIO.html

Schultz, R., 1999, Application of Strict Product Liability to Aeronautical Chart Publishers: Journal of Air Law and Commerce, v. 64, no. 2, p. 431-460.

Tarter, B., 1992, Information Liability: New Interpretations for the Electronic Age: Computer/Law Journal, v. XI, p. 484

Turner, P., and Groom, A., 2003, Negligence Law Update: Reform of the Law of Negligence: www.jws.com.au/pdf/175%20-%20Negligence%20Law%20-%20July%202003.pdf

West Publishing Company, 1998, West's Encyclopedia of American Law, St Paul MN, West Publishing Company, 5,500 p.

From CARIS GEMM 4 to *Carta for Geology* Software: Our Responses to Geology Educational Challenges in the Age of Digital Mapping

By H. Wouter van de Poll and Chris Parsons

CARIS
264 Rookwood Ave.
Fredericton N.B. Canada E3B 2M2.
Telephone (506) 458-8533
Fax (506) 459-3849 (CARIS)
E-mail vdp@caris.com, chris.parsons@caris.com

ABSTRACT

Geologists have been slow to adopt digital technology in their geological mapping. Too few geology departments offer digital geological mapping as part of their regular undergraduate/ graduate program. Many of the departments that do make computers available to their students for geological map-making have neither the software nor properly trained geology faculty members to teach the subject.

Reasons given by geologists for this lack of interest are many and varied. They range from costs, inappropriate software, and software that is too difficult to learn, to lack of time to learn a new technology. This situation does not bode well for any Web-enabled geological data distribution initiatives, which must rely on data-support and use by the geological community.

About 1989, some members of the faculty of the Department of Geology, University of New Brunswick, Fredericton, New Brunswick, Canada, decided that their present and future geology graduates should be given a chance to prepare themselves for the adoption of computer based geological mapping. To achieve this, they first developed a software prototype plus courseware, designed specifically for digital geological mapping, that can be taught at the community college/ university as well as the professional level; and second, they arranged to have this software commercially supported, for marketing and sales, maintenance, and further development according to industrial standards.

The resulting software was the **GE**ological **M**apping **M**odule (GEMM version 1, currently GEMM 4). The commercial support is by **CARIS**, a geographic information system (GIS) and hydrographic mapping software company specializing in the development of software applications for specific market niches, including geology.

This paper highlights what we see as essential factors in overcoming the reticence of geologists to adopt digital mapping technology in their geological studies. We will trace some of the history, developmental philosophy, and objectives of the GEMM 4, emphasizing its range of geological map-making capabilities. The current state of development of our new "Geological Map Processing System" software (provisional development code-name "*Carta For Geology*") will also be discussed.

INTRODUCTION

Success of initiatives such as the Geoscience Network (GEON) for regional or national distribution systems of earth science data depends on geologists' supplying quality, up-to-date, digital geological data. However, the required geological infrastructure for such support is not in place and there is considerable doubt about the ability of practicing geologists to supply digital map data in quantity.

In the world of GIS-based digital geological mapping the database tends to take precedence over the geological map in the acquisition, interpretation, and distribution of geological data. Historically, however, the hard-copy geological map, not the geological data set, has been (and in the mind of most geologists, still is) the preferred means for the interpretation, display, and exchange of geological data. In the face of this map-versus-data-set debate and for a variety of other reasons, geologists have remained largely apathetic with respect to digital geological mapping. As a result, much of the development of digital geological mapping software is still in an experimental stage.

Some commonly cited reasons for the slow adoption of digital mapping include:

- I don't like GIS, it forces me to change work habits and procedures.
- GIS is for geographers; it's not suitable for geological mapping.
- The time it takes to learn GIS affects my research

and publication record, hence my ability to attract research funding.
- Commonly used GIS software must be customized to my needs as a geologist. I do not have access to the technical support needed to accomplish that.
- The software packages that are out there cannot "talk" to each other, and isolate people into individual software camps.
- Computers are unreliable and the software is too difficult to learn.
- Going digital adds unnecessary expenses to my research.
- Geologists should be out there looking at rocks, not computer screens.
- The software limits my ability to display the data as an integrated geological map.

Commercial participation in the development of geological mapping software is slow. Universally adopted standards for the display and exchange of geological maps do not exist. As long as the market remains poorly defined and underdeveloped there is little incentive for software companies to become involved. Without commercial participation, innovation, and competition the development of mapping software specifically for geology is not likely to advance much beyond its current level.

Often, the digital technology used by geology students is centered on the use of graphics software rather than geo-referencing mapping software. Far too few geology departments teach digital geological mapping as part of their geology undergraduate/graduate program. Universities that do offer exposure to GIS do so mainly via their geography departments at the insistence of the geographers, not via geology curriculums. As a result, few geology faculty members, including newly hired junior faculty, have acquired the necessary skills and interests to initiate a suitable digital geological mapping program in their departments.

The geological application of whatever software is in use is usually left to the student and user to figure out. Is it any wonder that so many new geology graduates enter the profession with a stated dislike for digital geological mapping?

Clearly lost in the debate is that a computer equipped with appropriate software offers improved efficiency in the production and maintenance of geological maps and enhances the integrity of geology as a science by:

- using GPS technology for improved map accuracy;
- integrating map and associated data sets;
- centralizing data storage and web distribution, integration, analysis and interpretation of earth science data;
- speeding up the map release date; and
- allowing frequent map updates and colored map displays

Most importantly, going digital in field data capture, map preparation, and database creation:

- eliminates loss of original data (field notes);
- saves time, and hence, money; and
- offers the geologist maximum participation in the quality, flow, use, application, and destination of his/her geological data.

The lack of expansion of digital mapping in applied geology is not likely to change until the geological community makes a serious commitment to the adoption of computer technology. This requires that geological funding agencies, professional geologists' associations, educational institutions, and practicing geologists combine their efforts in:

- formulating geological objectives and specifications in the design and adoption of geological mapping software; and
- allowing this software to be taught by geology departments independently from other departments with geomatics interests.

Only then may the geological community be able to attract a cross-section of industrial partners willing to invest in the development of mapping software that geologists will use.

"Determining and interpreting three-dimensional relations between geological objects is one of the skills the geological profession has highlighted and pursued with vigor and success throughout its history. The subtlety and utility of this skill may elude those who do not understand its profound power in interpreting earth's dynamic processes and the chronological order in which they occurred. It is, however, one of the fundamental discriminators of our profession and as such defines the purpose of the geological map. *It is very important that this facet of geological work not be excluded from the database-intensive geological world we are about to create"* (adapted after Bob Ilchik, 2003, personal communication). We think it essential that in the next generation of geological mapping software this issue of how to generate a geological map and associated database be resolved to the satisfaction of both the geologist's and data manager's interests.

This map-versus-database issue must also be addressed at the educational level. If the geological database, not the geological map, becomes the main focus in the acquisition and dissemination of geological data then how are university geology departments to respond to this challenge? What are the consequences? Can we really learn to make and understand geological observations outside the spatial context of the geological map? If the data are not to be collected and displayed in their spatial context, how will the quality of the data be assessed and by whom? If geological mapping is no longer the central focus in the way we display the results of studying the earth, will geology departments continue to be able to attract quality students? Will geology departments continue to exist? What will their purpose and mandate be? Most geology faculty members would probably say "no" to these questions. Their likely response would be to ignore digital mapping altogether and continue to teach geology with a focus on the hard-copy geological map.

Our philosophy and approach to overcoming this problem is to recognize the challenge and develop geological mapping software, such as *Carta For Geology*, that addresses both the map and the database side of the issue equally, by automatically generating an attribute list for each feature displayed on the map. Furthermore, the new software will be able to display a properly symbolized geological map on the web in a range of native formats for viewing and querying (CARIS Spatial Fusion). These two main features should help maintain the geological map in its rightful place next to the database in professional geology and education.

ORIGIN AND FUTURE OF THE CARIS GEMM

CARIS is a company known worldwide for its hydrographic charting and GIS software. It specializes in the development of software applications for specific market niches. The CARIS GEMM, *Carta for Geology*, and CARIS Law Of The Sea (CARIS LOTS) software are three software applications centered on the earth sciences.

The GEMM was initially developed in 1989 by members of the faculty and staff at the University of New Brunswick Department of Geology, who have over 80 years combined experience in bedrock geological mapping, mineral exploration, and teaching.

The GEMM was meant to be taught as a regular undergraduate/graduate geology course and comes equipped with associated courseware. Its purpose was to provide geology students with a means of bringing their geological mapping output into the geo-referenced digital age *without affecting well established and accepted conventional mapping procedures*. The course has been taught as part of our field school program for the past decade.

CARIS' roots in mapmaking, its production-oriented design, and commitment to backward compatibility encouraged us to adopt CARIS as a base for the GEMM. In 1993, the prototype of our GEMM was transferred to CARIS for continued development and support. As a result, the GEMM 4 is one of the few commercially available software applications specifically designed for geological work. It allows the geologist to efficiently complete, during the course of fieldwork, a preliminary digital geological map, GIS-ready and in full color. GEMM-based geological mapping will typically reduce the time period between the last day of fieldwork and the release date of the preliminary geological map from as much as 14 months to a few hours.

The GEMM has been continuously upgraded since its inception. By placing the emphasis on spatial thinking in the acquisition of geological data, efficient map production and overlay analysis (three aspects most familiar to practicing field geologists) the GEMM is meant to respond to most of the reservations held by field geologists about digital mapping.

The next generation GEMM *Carta for Geology*, under development now, is being redesigned to better address the needs of modern earth science data storage, web distribution, and database analysis without sacrificing well established geological mapping techniques.

GEMM 4 Installation

Included with the GEMM 4 installation procedures are:

- selection of either 'strike-dip' or 'dip direction-dip;"
- full CARIS documentation;
- specialized "self-teach" GEMM 4 courseware, consisting of twelve tutorials necessary for the successful completion of a geological mapping project; and;
- a complete set of training maps and files.

SELECTED COMPONENTS OF THE GEMM

The GEMM 4 Program Bar contains the following programs:

CARIS Tools. This facility consists of some 105 programs in 16 modules, covering a complete range of map-making, import and export, topology, and map editing executables.

Help Utility. Help files offer assistance in CARIS procedures, including geological assistance.

User Guides and Tutorials. Tutorials included with the software are centered on self-teaching CARIS, including self-help courseware in support of geology.

Plot Composer. This program is designed to accept and combine (OLE)-compliant objects, photos, logos, graphs, spreadsheets, etc., with GEMM-map metafiles for creating paper map or poster products.

***CARIS GEMM 4 Editor* (Data Entry).** This interactive program allows the creation and editing of digital maps. It is a version of the CARIS Editor customized for geological work. The CARIS courseware is centered on the efficient application of the GEMM 4 Editor.

Geological data entry is by three principal means:

1) The Geological Menu Bar includes a Geology menu tree supporting the "drag-and-drop" procedure for placing symbols including Oriented Annotated Symbols (strike-and-dip symbols, strike-and-plunge symbols) for efficient "free-hand" or "registered" map editing (fig. 1). CARIS layers can be selected on the basis of:

- Feature Codes, User Numbers, Theme Numbers, Keys or Data Type.

Geological Menu items include Wizard panels for:

- Creating a new CARIS map (Header);
- Importing foreign data formats and imagery;
- Merging and extracting maps;
- Batch data input;
- Creating regular and irregular digital terrain models;
- Adding pre-defined scale-dependent borders and/or scale bars;
- Automated network and polygon topology building;
- Many other useful geological map-making functions; and
- Geological Help files.

2) A Digitizer Geological Template Menu for easy and efficient digitization of geological line and point symbols. The digitizer Geological Template Menu allows the user to digitize geological symbols and line types from registered paper maps, field slips, aerial photographs, etc. onto the screen-map. Layer Tags are pre-set when digitizing from the template menu (fig. 2).

3) A GPS Field Data Collection Utility for locational accuracy and database entry. The CARIS GPS field data collection utility provides a direct link to the Global Positioning System (GPS) collection unit; allowing the user to download waypoints while simultaneously populating any Open Database Connectivity Compliant (ODBC)-attribute database. The waypoints are inserted into the map/database file using pre-selected geological symbology. The user interface is designed to automatically respond to the type of data to be collected. The utility displays a geotiff image of the map on which the data collected in the field are shown. The database interface within the map file consists of customizable forms according to the application of the user. They are accessible by clicking on the waypoint.

The GEMM courseware covers the following five steps commonly involved in completing a CARIS-based geological-mapping project:

1) Building a project basemap either by:

- hand digitization *or*
- on-screen Semi Automated Map Input (SAMI), also referred to as "heads-up" digitization (fig. 3).
- By map import from a range of common formats including E00, Shapefile, SDTS, DXF, etc. (fig. 4).
- Or by combining elevation contours obtained from imported digital elevation (ASCII X, Y, Z) data, with imported digital line data (see above).

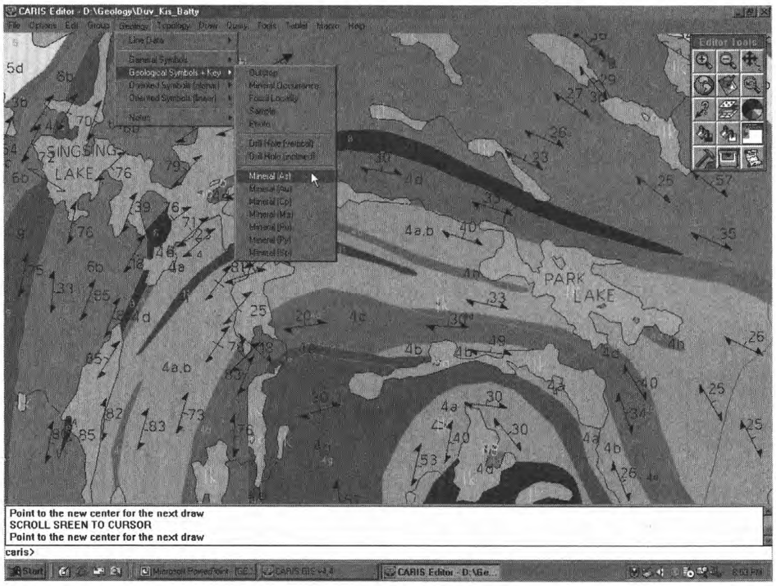
2) Adding the geology consisting of digitized line and point symbols using the GEMM Template in combination with registered field slips and/or aerial photographs.

3) Building network and polygon topology includes semi-automated map editing and supports the execution of a common range of GIS analytical procedures.

4) Building the legend so that the geological map can be distributed as a preliminary finished product in hard copy and/or digital format.

5) Adding polygon colors to the map file for improved viewing and map interpretation.

The way these five steps are approached may differ



```
Point to the new center for the next draw
SCROLL SREEN TO CURSOR
Point to the new center for the next draw
caris>
```

Figure 1. Example of a GEMM geological map displaying the GEMM 4 menu bar and Geology menu tree.

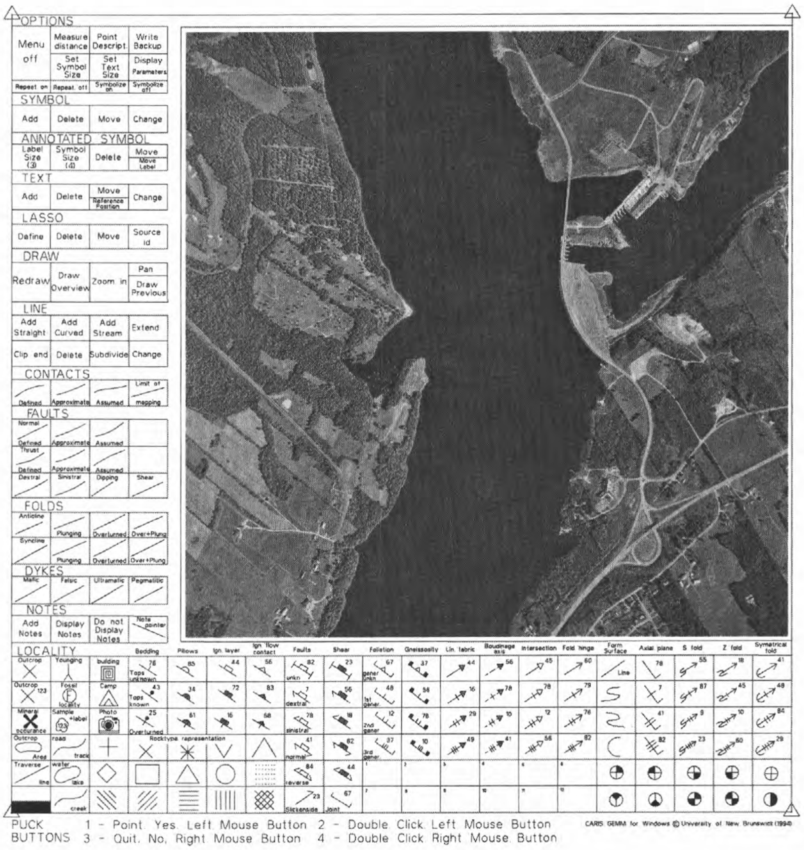


Figure 2. The Digitizer Geological Template menu.

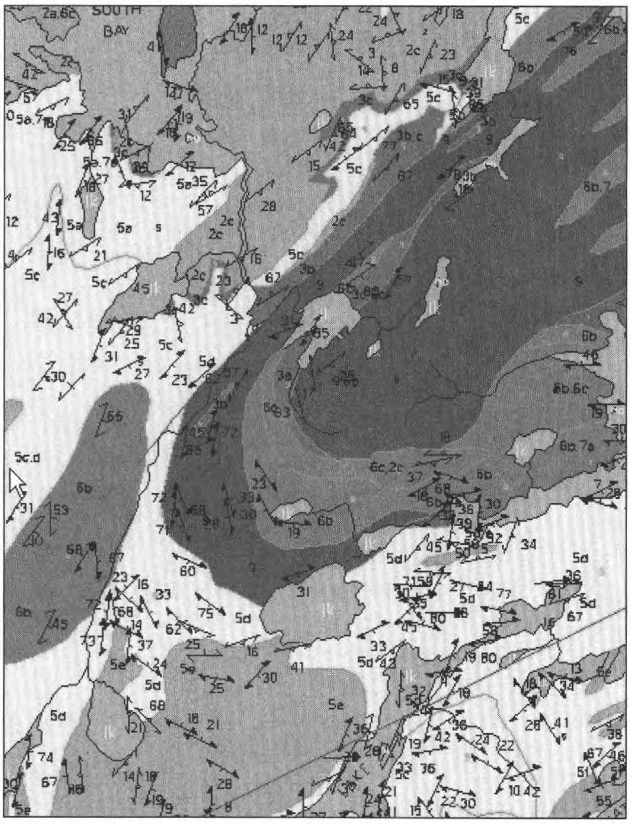


Figure 3. Using CARIS SAMI, a scanned geological map can be converted into a digital map file retaining all its geology symbology.

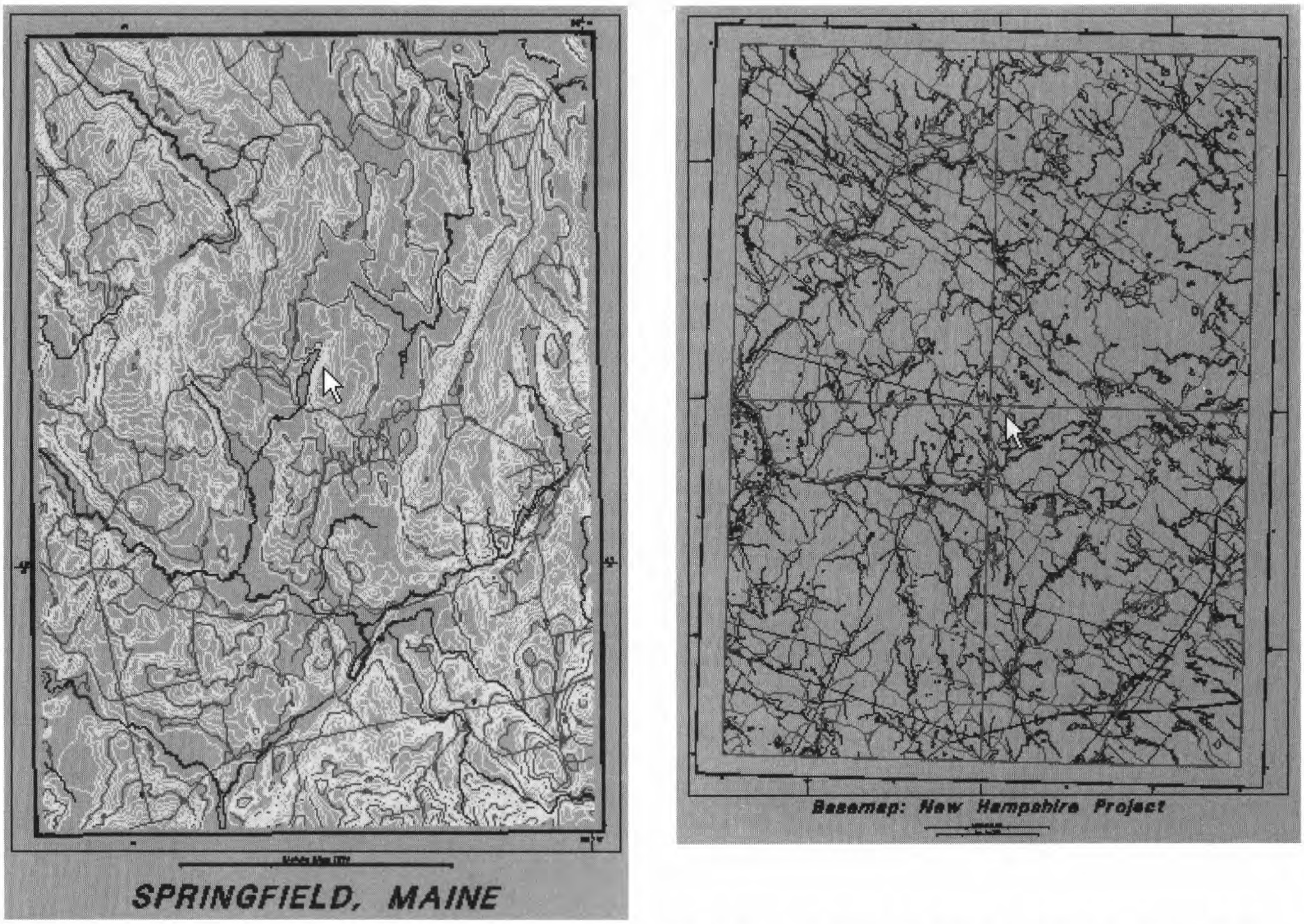


Figure 4. Topographic basemaps created for Springfield, Maine (E00 Format) and part of New Hampshire (SDTS Format) imported into CARIS GEMM 4.

from project to project. They can be accommodated, however, at any level of complexity.

CARIS Manager with Zones. The CARIS Information Manager allows the management and querying of geographic and attribute data residing in a Geographic Information Database.

Digital Terrain Model (DTM) Viewer. The GEMM 4 provides DTM functionality for both triangulated irregular network (TIN) and regular gridded data sets. The DTM functionality includes a 3-dimensional (3-D) viewer, with adjustable viewing angle, sun elevation, adjustable projection, vertical exaggeration and other display settings.

The user can also create contours and profiles, calculate volumes, and classify the DTM by elevation, slope or aspect. Tools are provided for draping map and vegetation features over a DTM, allowing the user to evaluate the effects of industrial development (such as open-pit mines) in the context of esthetic objectives and concerns.

Raster Imagery. The DTM function also provides for shaded black and white as well as color raster images from any X, Y, Z data using a gridded DTM executable.

When generating shaded-relief raster images, the resolution, sun direction and elevation angles can be defined by the user to assist in geological lineament analysis. The GEMM 4 allows pixels within raster images to display both color, raw, and calculated data values, providing a basis for accurate data-recall and overlay analysis.

File Management and Multi-media. This program supports hot linking to various multi-media sources, allowing for the display of photos, drill logs, scanned sketch maps and other media, simply by clicking on a feature on the digital map.

Database Manager. Database connections with any ODBC database, allowing the importation of geo-spatial data directly onto the user's digital map, are covered in the Database Manager. This program allows the user to perform table joins and multiple database queries, using the industry standard one-to-many protocol. It is being redesigned to more specifically respond to current geology database issues and objectives.

GEMM 4 EVALUATION

GEMM 4 provides for a range of geological data-entry, map creation, display and overlay analytical procedures that should satisfy most of the immediate needs of practicing field geologists (fig. 5).

Nevertheless, we recognize that improvements can be made to the geological database aspect of GEMM so that generating a database becomes an integral part of the mapping process. We also see a need to improve the way GEMM supports statistical functionalities and allows querying of attribute data. Furthermore, GEMM's appearance is becoming old fashioned and legend building is more time-consuming than is desirable.

The evolution of GEMM is an ongoing process. Improvements to GEMM 4 will be addressed in our next generation of geological mapping software, *Carta for Geology* (provisional name).

BRIEF OVERVIEW OF *CARTA FOR GEOLOGY*

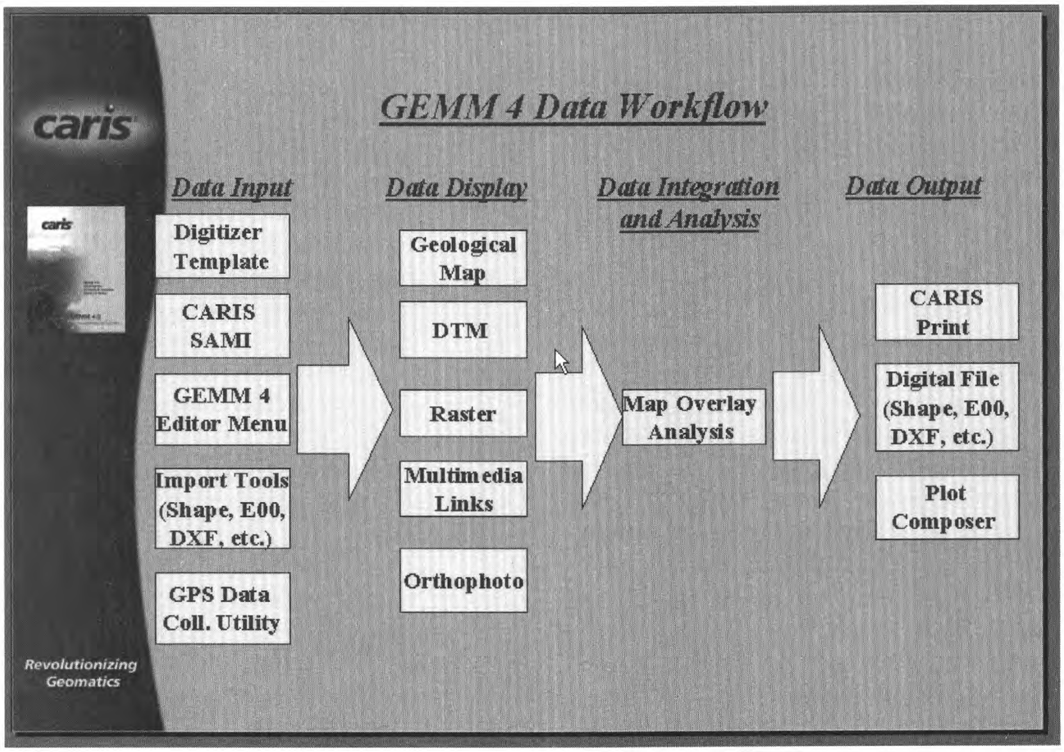
With respect to GEMM 4, *Carta for Geology* (fig. 6) is designed to better handle user-defined objects and other aspects related to modern geology independent of where and how they are stored. Attributes for these objects can be stored in a number of standard formats including Oracle, Access, CARIS GEMM files and shapefiles. Like GEMM 4, *Carta for Geology* will support geological symbology.

Some of *Carta for Geology*'s current main features are:

- A modern user interface with dockable windows.
- Direct map-database linkage, automatically creating a map-generated, ODBC database.
- Geological statistics (stereonet plots).
- Ability to support shapefiles natively.
- Editing and digitization including raster to vector map input.
- Annotation.
- Query builder.
- Plug-ins and customization.

Major *Carta for Geology* features to be completed in the near future include:

- Ability to edit shapefiles.
- Data Dictionary and data mapping.
- Expanded statistics
- Ability to support any datamodel.
- Semi-automated Legend builder.
- Gridding package.
- Improved 3-D visualization and analysis.



caris

GEMM 4 Data Workflow

Data Input

Digitizer Template

CARIS SAMI

GEMM 4 Editor Menu

Import Tools (Shape, E00, DXF, etc.)

GPS Data Coll. Utility

Data Display

Geological Map

DTM

Raster

Multimedia Links

Orthophoto

Data Integration and Analysis

Map Overlay Analysis

Data Output

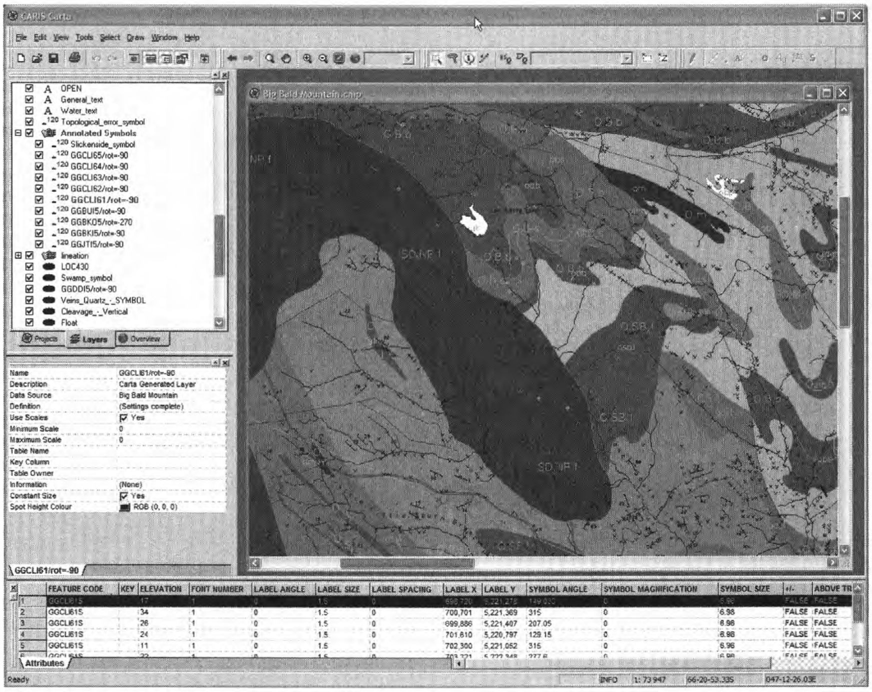
CARIS Print

Digital File (Shape, E00, DXF, etc.)

Plot Composer

Revolutionizing Geomatics

Figure 5. Flow diagram for the GEMM 4.



☑ A OPEN
☑ A General_text
☑ A Water_text
☑ 120 Topological_error_symbol
⊟ ☑ Annotated Symbols
  ☑ 120 Slickenside_symbol
  ☑ 120 GGCLI65/rot=-90
  ☑ 120 GGCLI64/rot=-90
  ☑ 120 GGCLI63/rot=-90
  ☑ 120 GGCLI62/rot=-90
  ☑ 120 GGCLI61/rot=-90
  ☑ 120 GGBUI5/rot=-90
  ☑ 120 GGBKO5/rot=-270
  ☑ 120 GGBKI5/rot=-90
  ☑ 120 GGJTI5/rot=-90
⊞ ☑ lineation
☑ LOC430
☑ Swamp_symbol
☑ GGDDI5/rot=-90
☑ Veins_Quartz_-_SYMBOL
☑ Cleavage_-_Vertical
☑ Float

Name	GGCLI61/rot=-90
Description	Carta Generated Layer
Data Source	Big Bald Mountain
Definition	(Settings complete)
Use Scales	☑ Yes
Minimum Scale	0
Maximum Scale	0
Table Name	
Key Column	
Table Owner	
Information	(None)
Constant Size	☑ Yes
Spot Height Colour	■ RGB (0, 0, 0)

GGCLI61/rot=-90

	FEATURE CODE	KEY	ELEVATION	FONT NUMBER	LABEL ANGLE	LABEL SIZE	LABEL SPACING	LABEL X	LABEL Y	SYMBOL ANGLE	SYMBOL MAGNIFICATION	SYMBOL SIZE	+/-	ABOVE TR
1	GGCLI61S		17	1	0	1.5	0	698,720	5,221,278	149.030	0	6.98	FALSE	FALSE
2	GGCLI61S		34	1	0	1.5	0	700,701	5,221,369	315	0	6.98	FALSE	FALSE
3	GGCLI61S		26	1	0	1.5	0	699,886	5,221,407	207.05	0	6.96	FALSE	FALSE
4	GGCLI61S		24	1	0	1.5	0	701,610	5,220,797	129.15	0	6.98	FALSE	FALSE
5	GGCLI61S		11	1	0	1.5	0	702,300	5,221,052	315	0	6.96	FALSE	FALSE

Figure 6. Screen-view of *Carta for Geology* (Overview Map not shown) In the top left corner is the Control window displaying the Layer list of each item on the map. Here one organizes and modifies a project, selects the features to be drawn etc. Below it is the Properties window where one can edit the color, relative size, and scale ranges etc. of map features. The Attribute data window, displaying the attributes of selected map features, is positioned across the bottom. The geological map, or a zoomed in version of it, is shown in the Graphics window.

- Publish to the Web, using CARIS Spatial Fusion technology. Spatial Fusion is an Open GIS Consortium (OGC)-compliant data service capable of supporting a host of data formats that allow users to view and query your spatial data holdings over the Internet.

Carta for Geology is being specifically designed to respond to what we think are the current and future demands of modern Geology.

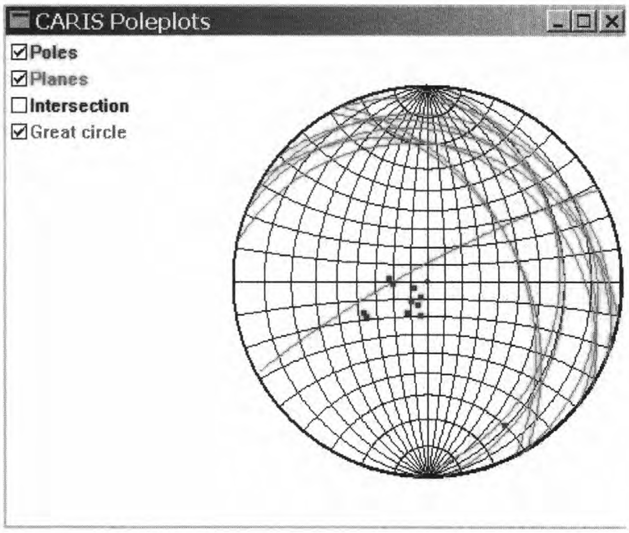
With respect to the future of digital geological mapping, it seems clear to us that *GIS has overwhelmingly failed to capture geologists' interest* to the extent that we no longer see emphasizing GIS as a viable way to market our geological mapping software. Accepting the premise that the main purpose of GIS is to support analysis, we suggest that the main purpose of geological mapping software should be efficiency in data acquisition, map display, map data analysis and the supply of data files *in support of GIS* in a range of common formats.

Except for the statistical display of structural data (fig. 7), geologists tend to be ambivalent about database issues; they see the database as detracting from the spatial aspects of geological mapping and filling out and maintaining database tables as not their responsibility.

However, being able to automatically populate and edit the Attribute Data List (database) by editing geological point and line data in their spatial context on the map, the way *Carta for Geology* does, should overcome this ambivalence. This map-database connectivity is specifically intended to bridge the gap between historical working methods and objectives of the mapping geologists and their priorities concerning the geological map, and the current objectives of the geology data managers and their priorities concerning database analyses.

Additional features that most applied geologists would likely be looking for in geological mapping software include:

- Maintaining the *properly symbolized* geological map as the premier scientific document in the interpretation, exchange and display of geological data.
- Choice in the selection of mapping software.
- Software that offers the geologist some control over the acquisition, flow, use, application and destination of his/her geological data.



☑ Poles
☑ Planes
☐ Intersection
☑ Great circle

Figure 7. Stereonet plot of a selection of first generation inclined cleavage orientations extracted from the *Carta for Geology* attribute list.

- A mapping system with GIS functionality that allows a proper geological map to serve as the basis for overlay analysis to which other data layers can be added according to the nature of their projects.
- Software they can learn to use with relative ease, and that is flexible and responsive to their needs.

CONCLUSIONS

Geologists are entering the world of digital mapping slowly and with a great deal of trepidation. This is understandable considering the geologists' widespread fear that GIS mapping objectives will force them to give up the geological map as the principal means for the display, interpretation, and dissemination of geological data in exchange for the geological database.

Current software widely used in geology makes it relatively easy to disseminate, query and analyze geological data in database format but does not meet the demands for easy map creation. The continued emphasis on this type of software in geological mapping can have long-term educational and professional implications that may seriously affect the quality and supply of digital geological maps.

Recognizing this potential problem, the intention of this paper is to draw attention to the need for geological mapping software that, while being capable of supporting GIS analysis, retains a strong geological focus as defined from the perspective of practicing geologists. The CARIS GEMM 4 and *Carta for Geology* are used here as examples of the type of software that may gain the confidence of the geological community. *Carta for Geology* in particular is designed to straddle the divide between proponents of the geological map and proponents of the geological database. In addition, it allows web-client access to both components for viewing and querying in a range of native formats using CARIS Spatial Fusion.

The National Geologic Map Database Project: Overview and Progress

By David R. Soller[1] and Thomas M. Berg[2]

[1]U.S. Geological Survey
926-A National Center
Reston, VA 20192
Telephone: (703) 648-6907
Fax: (703) 648-6977
e-mail: drsoller@usgs.gov

[2]Ohio Geological Survey
4383 Fountain Square Dr.
Columbus, OH 43224
Telephone: (614) 265-6988
Fax: (614) 268-3669
e-mail: thomas.berg@dnr.state.oh.us

The National Geologic Map Database (NGMDB) project continues to fulfill its mandate. Some of its accomplishments are specific and tangible, and others are more general in nature – for example, the NGMDB contributes to advancements in digital mapping techniques and database design by agencies in the United States and internationally. However, without extensive collaboration from enthusiastic and highly skilled members of the state geological surveys and the Geological Survey of Canada, these accomplishments would not have been possible. Highlights of the past year include:

- the Geoscience Map Catalog now contains bibliographic records for more than 61,000 map products published by more than 270 organizations including the U.S. Geological Survey(USGS), 43 state geological surveys, universities, and scientific societies and organizations,
- the Geologic Map Image Library has evolved from a concept to a prototype Web site that serves high-resolution images of nearly 1,000 geologic maps,
- the project contributed significantly to evolution of the North American standard data model, science language, and data-interchange format, and to the cartographic standard for the U.S. Through discussions with ESRI, this data model may form the basis for their Geology Data Model for Arc Geodatabase. Internationally, NGMDB staff participated in "DIMAS", the map standards committee of the Commission for the Geological Map of the World,
- the seventh annual Digital Mapping Techniques workshop was a success, bringing together 90 technical experts from 36 agencies, and
- the third phase of the project – the design and implementation of an online, vector-map database – was reoriented mid-year, and began to focus on data input tools and standardized science language.

INTRODUCTION

This project provides an unusual if not unique opportunity to foster better relations and technical collaboration among all geological surveys in the nation. Given the nature of the issue – the creation and management of geoscience map information in digital format during a period of rapid technological evolution – collaboration is critically important. Perhaps more significant, these are changing times for all geological surveys – funding and staff seem to become more scarce each year – and through collaboration we can share our intellectual and computing resources and not "reinvent the wheel" within each agency.

Before describing the NGMDB components and progress, we wish to highlight the various mechanisms by which we define and accomplish our goals. Because advice, guidance, and technical collaboration are an integral part of this project, we discuss the project plan at numerous venues throughout the year. These include geoscience and related professional society meetings, the Digital Mapping Techniques workshop, and site visits to state geological surveys. Advice gathered at these venues serves to refine and, in some cases, to redirect the project's goals. Comments from users, generally via our Web feedback form, also provide us with valuable perspectives, and have prompted us to make numerous modifications, especially to our Web interface design.

Because the NGMDB's scope is so broad, its success relies on the many people and agencies that participate in its activities. Members of the committees and small working groups that advise and contribute to the project's goals are listed in Appendix A. These committees are an important mechanism for coordinating with each agency, and they deserve noting:

- Digital Geologic Mapping Committee of the Association of American State Geologists (AASG) – charged with representing all state geological surveys in the NGMDB project, and with providing authoritative guidance to the project.
- Technical Advisory Committee – provides technical vision and guidance to the NGMDB, especially on the project's Phase Three.
- Map Symbol Standards Committee – oversees the completion, and then the maintenance, of the Geologic Map Symbolization Standard, which will become a Federal standard endorsed by the Federal Geographic Data Committee.
- AASG/USGS Data Capture Working Group – coordinates the annual Digital Mapping Techniques workshop, and provides through an email listserver a forum for exchange of technical information.
- AASG/USGS Metadata Working Group – summarized issues related to creating metadata, and identified useful software tools.
- AASG/USGS Data Information Exchange Working Group – created technical guidance for map publication guidelines.
- AASG/USGS Data Model Working Group – defined a draft version of a standard geologic map data model.
- North American Data Model Steering Committee – succeeded the Data Model Working Group, and is developing a standard data model, science language, and data-interchange format for the North American geoscience community.
- NGDMB contact-persons – within each state geological survey, several people work with us on various project databases and activities.

BACKGROUND

The National Geologic Mapping Act of 1992 and its reauthorizations in 1997 and 1999 (PL106-148) require a National Geologic Map Database to be built by the USGS in cooperation with the AASG. This database is intended to serve as a "national archive" of standardized geoscience information for addressing societal issues and improving our base of scientific knowledge. The Mapping Act anticipates a broad spectrum of users including private citizens, professional geologists, engineers, land-use planners, and government officials. The Act requires the NGMDB to include these geoscience themes: geology, geophysics, geochemistry, paleontology, and geochronology.

In mid-1995, the general stipulations in the Geologic Mapping Act were addressed in the proposed NGMDB design and implementation plan developed by the USGS and AASG. Summaries of this plan are listed in Appendix B. Because of the mandate's broad scope, we proposed a phased, incremental design for the NGMDB. A phased approach has two benefits: 1) it enables us to identify the nature and quality of existing information and quickly serve it to the public; and 2) it gives us time to build consensus and expertise among the database designers in the state geological surveys and the USGS. Furthermore, it enables us to more effectively consider and respond to evolving technology and user needs. These phases, and our progress, are shown in figure 1.

In the first and most fundamental phase of the project, we are building a set of easy-to-use reference databases; for example, a comprehensive, searchable map catalog of all geoscience maps in the United States, whether in paper or digital format. The second phase of the project focuses on the development of standards and guidelines needed to improve the utility of digital maps. The third phase proposes to, in the long term, develop an online database of vector-based geologic map information at various scales and resolution.

In late 1995, work began on Phase One. The formation in mid-1996 of several AASG/USGS Standards Working Groups initiated work on Phase Two. The project opened its Web site to the public in January, 1997, as a prototype intended to solicit comments on the Map Catalog. At the Digital Mapping Techniques '98 through '03 workshops, a series of presentations and discussion sessions provided updates on the NGMDB and, specifically, on the activities of the Standards Working Groups. These progress reports are listed in Appendix B. This report summarizes accomplishments since the project's inception, and therefore repeats material from previous reports, but it focuses on activities since mid-2002. Additional and more current information may be found at the NGMDB project-information Web site, at <http://ncgmp.usgs.gov/ngmdbproject>. The searchable databases are available at <http://ngmdb.usgs.gov>.

To submit general comments about project scope and direction, please address the authors directly. For technical comments on the databases or Web page design, please use our Web feedback form; this form is linked from many of our search pages (see "Your comments are welcome", at http://ngmdb.usgs.gov/).

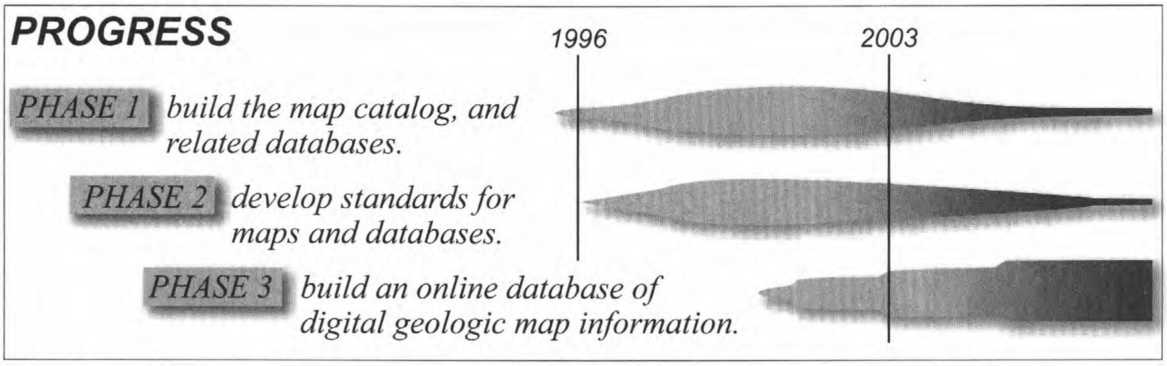


Figure 1. Diagram showing the three NGMDB Phases, and progress toward our goals (for example, documenting in the Geoscience Map Catalog all maps and related products for the United States and its territories and possessions).

PHASE ONE

Through ongoing discussions with private companies, citizens, government officials, and research geologists, it is clear that first and foremost, we need to provide reference databases so that geoscience maps and descriptive information can be found and used. Many people want to better understand the geologic framework beneath their home, business, or town, and so we are building several databases that support general, "data-discovery" questions posed by citizens and researchers alike (fig. 2). These reference databases are: 1) the Geoscience Map Catalog; 2) GEOLEX, the U.S. geologic names lexicon; 3) Geologic Mapping in Progress, which provides information for ongoing National Cooperative Geologic Mapping Program (NCGMP) mapping projects, prior to inclusion of their products in the Map Catalog; and 4) the prototype version of our Geologic Map Image Library – this new initiative is briefly described below, and in other papers in this volume. Plans for the prototype National Paleontology Database also are discussed below.

Figure 3 shows the number of people (actually, the number of unique IP addresses or computers) who have used the NGMDB, per month since it opened to the public in January, 1997. These numbers indicate that the site has become a useful resource. Additional increases in use are expected as the Map Catalog, Geolex, and Image Library become fully populated.

The Geoscience Map Catalog

"I want to know if a map exists for an area, and where I can find it…"

Many organizations produce paper and digital geoscience maps and related products. Discovering whether a product exists for an area, and if so, where it can be purchased or obtained online, can be a time-consuming process. In the past, people found this information by contacting various agencies and institutions, and by conducting extensive library searches. To increase accessibility and use of these paper and digital products, we built the Geoscience Map Catalog as a comprehensive, searchable database of all maps and related products for the United States and its territories and possessions.

The Geoscience Map Catalog contains bibliographic records for more than 61,000 products from at least 270 publishers (see Appendix C or our most current list of publishers at http://ngmsvr.wr.usgs.gov/ngmdb/pub_series.html). Most of these products are from the USGS and 43 state geological surveys. Other publishers include state agencies, federal agencies, scientific societies, park associations, universities, and private companies. Products range from digital maps to books that don't contain maps but describe the geology of an area, and can be formal series products, open-file reports, or unpublished dissertations (fig. 4). Because there are many types of geoscience maps and related products, we categorize them by theme (fig. 5).

The Geoscience Map Catalog provides links to more than 1,300 published, downloadable products of the USGS and the state geological surveys. These links are established only to stable Web pages that provide the official copy-of-record for the publication – in the USGS, links are established only to the Publications Server and the NSDI Clearinghouse node.

Figure 6 shows how the Geoscience Map Catalog can be used to find particular products – upon searching it and identifying the needed product(s), the user is linked to the downloadable data and metadata, to a depository library, or to the appropriate organization for information about how to purchase the product. We address the diverse needs of our user audience through four search options. The easy-to-use Place Name Search is based on the USGS Geographic Names Information System (GNIS); it is designed mostly to address the needs of non-geologists

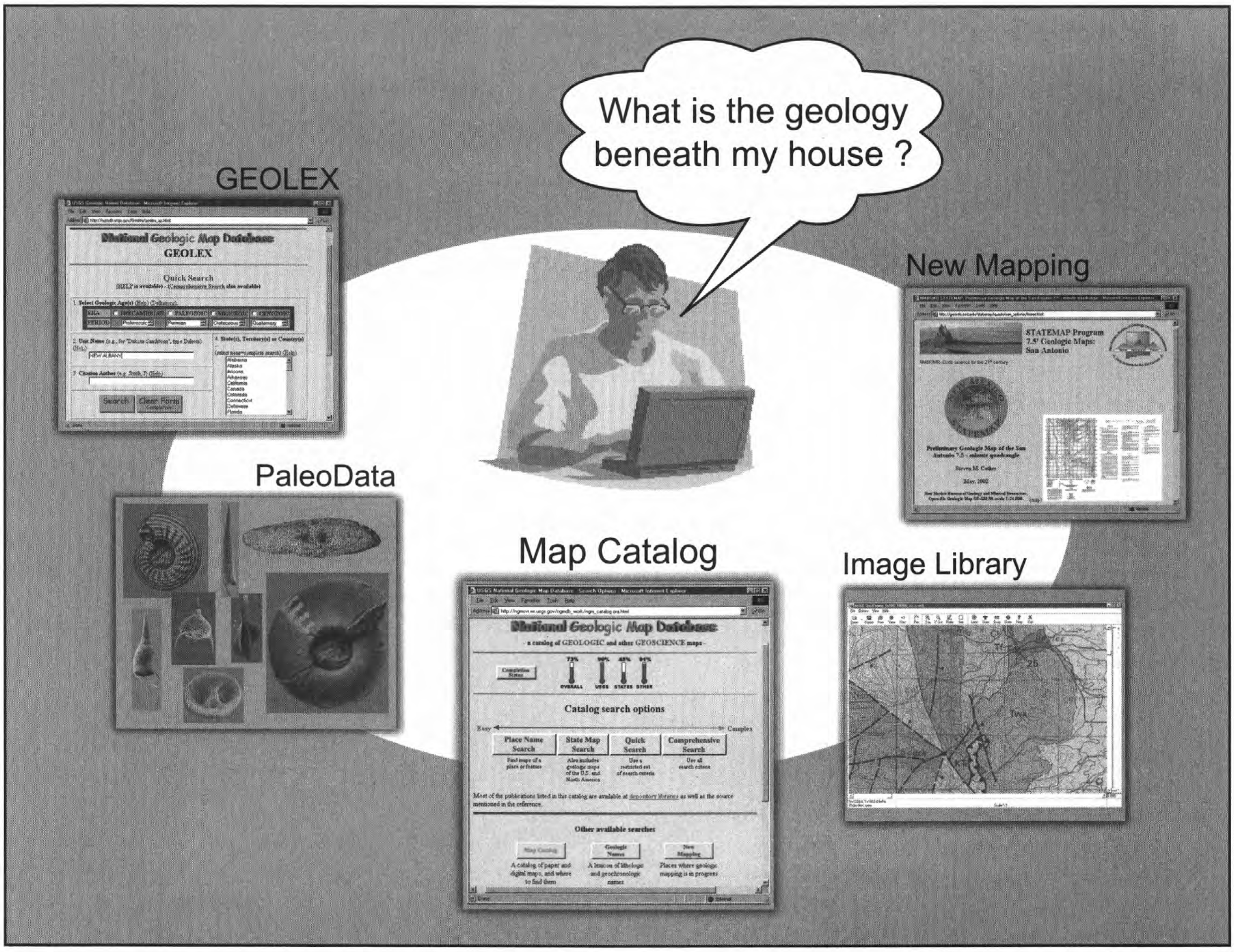


Figure 2. Many people want to know whether the geologic framework and the geoscience characteristics (for example, earthquake hazard, geochemistry) of an area have been studied and published. The reference databases built under NGMDB Phase One provide users with access to that information.

who want to use a simple interface to find information about their home, town, or worksite (fig. 7). In contrast, other choices such as the Comprehensive Search offer more search criteria.

The U.S. Geologic Names Lexicon ("GEOLEX") "I want to know more about the geologic units shown on this map…"

This is the nation's lexicon of geologic nomenclature. GEOLEX contains information for more than 16,000 geologic units in the U.S. (Stamm and others, 2000). It is an excellent resource for finding significant publications that defined and described geologic units mapped in the U.S. These publications can be critically important in field studies, enabling students and mappers to compare these published descriptions with what they see in the field.

GEOLEX includes the content of the four geologic names databases on USGS Digital Data Series DDS-6 (Mac Lachlan and others, 1996). Before incorporating into GEOLEX, those databases were consolidated, revised, and error-corrected. Our work now focuses on:

1. resolving the name conflicts found in the four databases of Mac Lachlan and others (1996). This is done by consulting publications, previous U.S. geologic names lexicons (listed in Appendix A of Stamm and others, 2000), and the records of the U.S. Geologic Names Committee (GNC),
2. using the previous lexicons to incorporate type locality, publication history, geologic age, areal extent, and usage information for many central and western U.S. geologic units listed in Mac Lachlan and others (1996),
3. adding geologic names not recorded in Mac Lachlan and others (1996) but found in the old USGS regional geologic names card catalogs (this is estimated to be 25% of all U.S. names), and

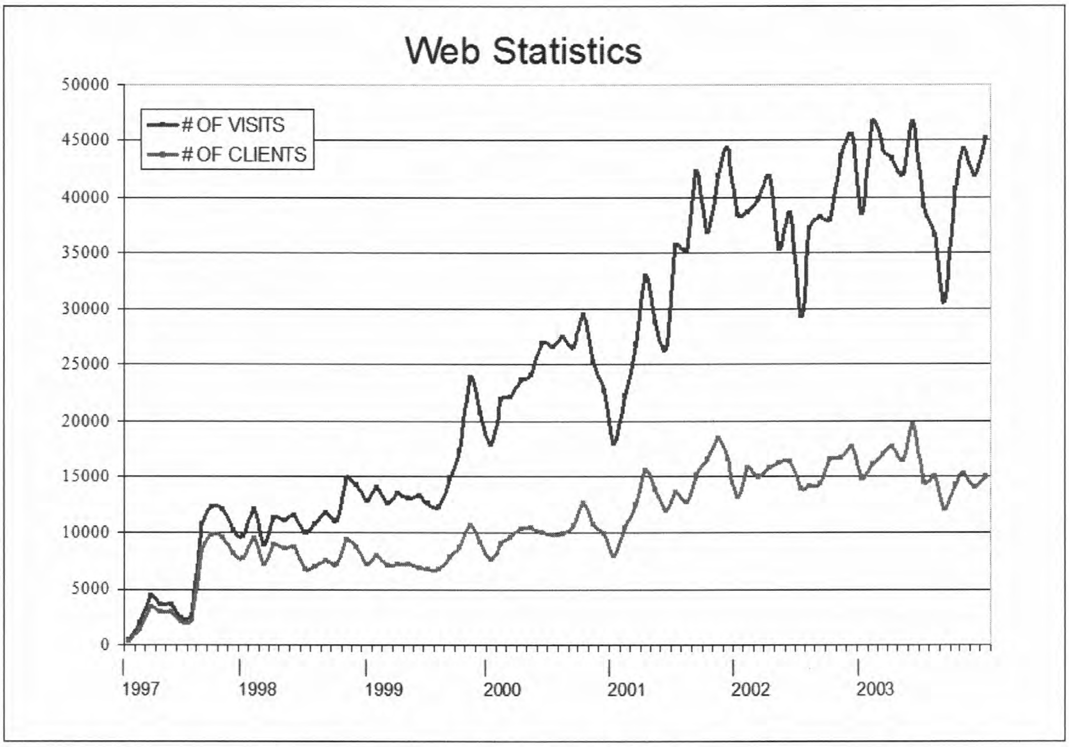
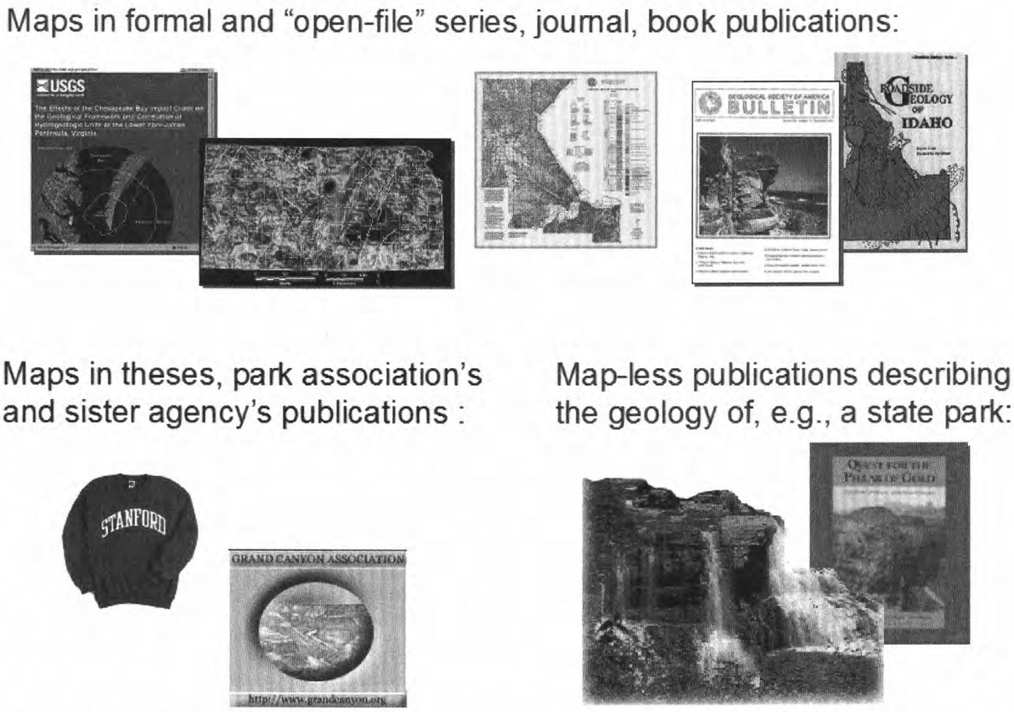


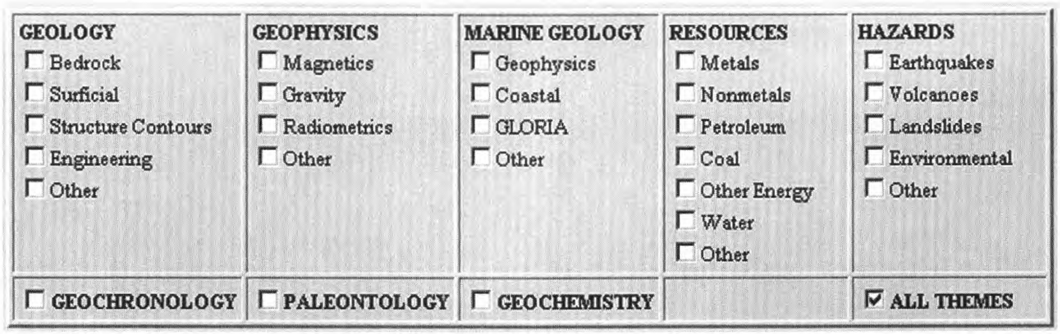
Figure 3. Web usage for the Geoscience Map Catalog, GEOLEX, and Mapping in Progress Databases. This diagram shows that the number of people (actually, the number of unique IP addresses or computers) using the NGMDB has gradually increased as these resource databases become more widely known; this usage trend is punctuated by sharp increases after essentially all USGS maps were entered into the Catalog and after many state geological surveys began to enter map records. The Catalog accounts for about 75-80% of user visits to the NGMDB site.


Maps in formal and "open-file" series, journal, book publications:

Maps in theses, park association's and sister agency's publications :

Map-less publications describing the geology of, e.g., a state park:

Figure 4. Bibliographic records in the Geoscience Map Catalog are drawn from a diverse group of more than 270 publishers.



GEOLOGY	GEOPHYSICS	MARINE GEOLOGY	RESOURCES	HAZARDS
☐ Bedrock	☐ Magnetics	☐ Geophysics	☐ Metals	☐ Earthquakes
☐ Surficial	☐ Gravity	☐ Coastal	☐ Nonmetals	☐ Volcanoes
☐ Structure Contours	☐ Radiometrics	☐ GLORIA	☐ Petroleum	☐ Landslides
☐ Engineering	☐ Other	☐ Other	☐ Coal	☐ Environmental
☐ Other			☐ Other Energy	☐ Other
			☐ Water	
			☐ Other	
☐ GEOCHRONOLOGY	☐ PALEONTOLOGY	☐ GEOCHEMISTRY		☑ ALL THEMES

Figure 5. A portion of the Geoscience Map Catalog search page, showing the types of products included.

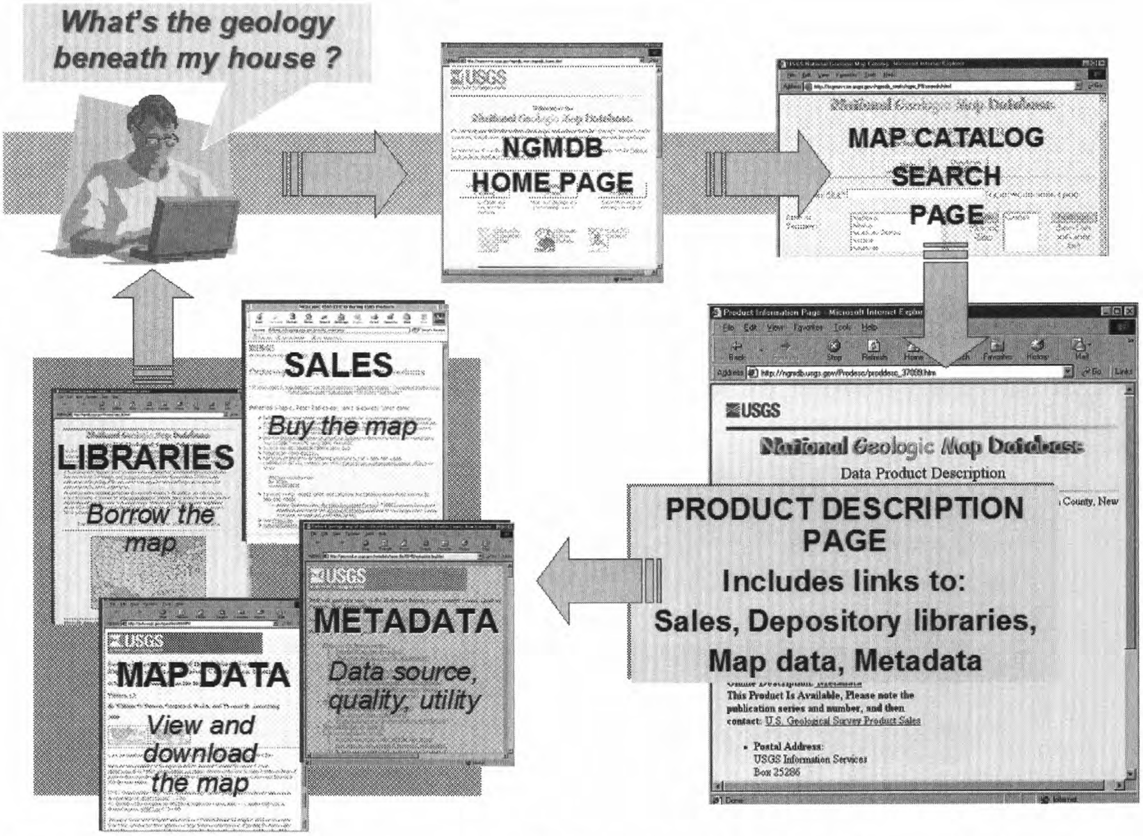
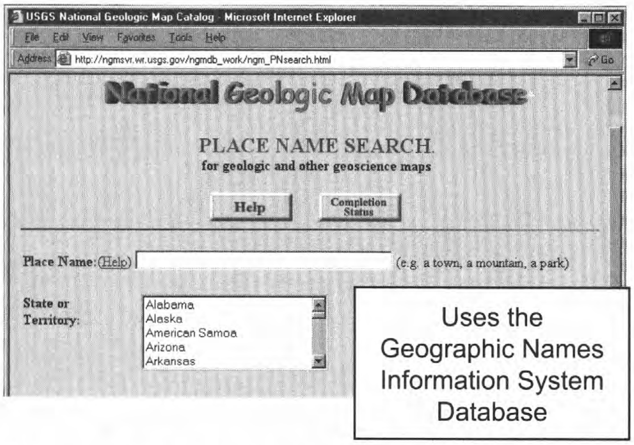


Figure 6. Diagram showing how a user navigates the Geoscience Map Catalog. Interested in knowing something about the geology of an area (such as the land beneath their house), the user queries the Catalog, which returns a hit list of possibly useful maps and related products. The user selects one of these and, from the Product Description Page, obtains further information and can then choose to buy the product, view and download it, inspect the metadata, or find it at a depository library.



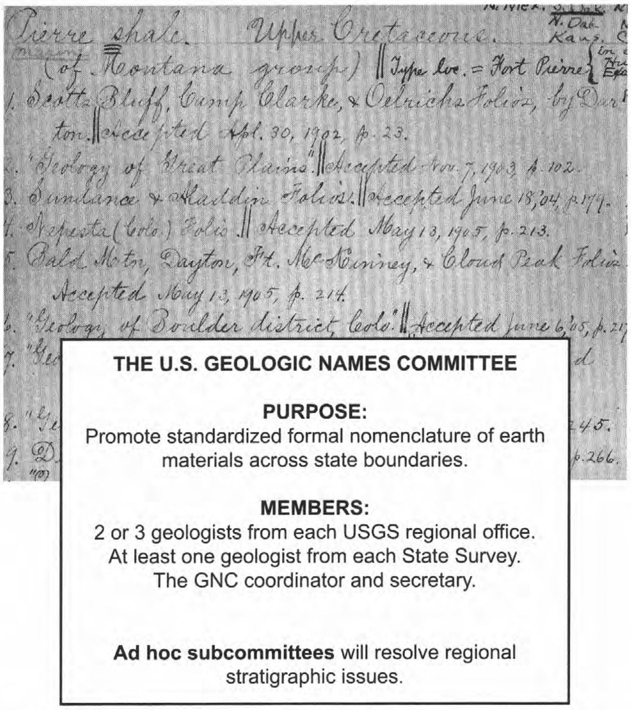
PLACE NAME SEARCH
for geologic and other geoscience maps

Place Name: (Help) (e.g. a town, a mountain, a park)

State or Territory:

Uses the Geographic Names Information System Database

Figure 7. The first page of the Geoscience Map Catalog's Place Name Search.



Pierre shale. Upper Cretaceous. N. Mex. N. Dak. Kans.
(of Montana group) || Type loc. = Fort Pierre
1. Scotts Bluff, Camp Clarke, & Oelrichs Folios, by Darton. || Accepted Apl. 30, 1902, p. 23.
2. "Geology of Great Plains" || Accepted Nov. 7, 1903, p. 102.
3. Sundance & Aladdin Folios. || Accepted June 18, '04, p. 177.
4. Nepesta (Colo.) Folio || Accepted May 13, 1905, p. 213.
5. Bald Mtn, Dayton, Ft. McKinney, & Cloud Peak Folios.
Accepted May 13, 1905, p. 214.
6. "Geology of Boulder district, Colo." || Accepted June 6, '05, p. 21
7. "Geo d
8. "Ge 45.
9. D p. 266.

THE U.S. GEOLOGIC NAMES COMMITTEE

PURPOSE:
Promote standardized formal nomenclature of earth materials across state boundaries.

MEMBERS:
2 or 3 geologists from each USGS regional office.
At least one geologist from each State Survey.
The GNC coordinator and secretary.

Ad hoc subcommittees will resolve regional stratigraphic issues.

Figure 8. The purpose and membership of the reconstituted Geologic Names Committee. Background image is an index card from the files of the USGS Geologic Names Committee, ca. 1903, showing decisions recorded regarding the use of the Pierre Shale in the USGS Geologic Atlas of the United States folios.

4. adding geologic names approved by the state geological surveys but not recorded in GEOLEX.

Many state geological surveys have been registering new geologic names with the USGS for decades, and are encouraged to continue this practice. In order to promote standardized geologic nomenclature within the U.S., the GNC is being reconstituted. Formerly a committee that focused on nomenclature issues within the USGS, the new GNC will include members from each state geological survey (fig. 8). When a conflict arises, GNC members from the USGS and those states affected will resolve it, and any changes will be recorded in GEOLEX. Through this mechanism, we anticipate that GEOLEX will serve the entire U.S. geoscience community.

Geologic Mapping in Progress Database

"I see from the Map Catalog that a map hasn't been published for this area – is anyone mapping there now?"

Our Geologic Mapping in Progress Database provides users with information about current mapping activities (mostly at 1:24,000- and 1:100,000-scale, but at 1:63,360- and 1:250,000-scale in Alaska) that is funded by the National Cooperative Geologic Mapping Program. We are re-engineering and repopulating this database, and will be linking it directly to the state geological survey fact sheets and Web sites.

Geologic Map Image Library

"I want to see a picture of this geologic map, online…"

Through discussions with users, and from comments received via our Web feedback form, it became clear that many people are interested in viewing and/or obtaining maps "online." Interpretation of the phrase "providing maps online" varies widely—to some people, it implies access to fully attributed, vector-based map databases, whereas to other people, it implies access to map images. Regarding the vector-based map database, we address this large task in Phase Three, below. With the Image Library, we have begun to provide map images to users, as described in two papers in this volume. We hope this new initiative will further strengthen the cooperative relationship between the AASG and USGS.

Paleontology Database

"I want to know if there is any fossil data from this area…"

The NGMDB project has designed and is planning to develop a National Paleontology Database (see Wardlaw and others, 2001). Our general plan is to build prototypes of this database in areas where geologic mapping is underway, so that we can work with mapping projects to design a database useful to science as well as to the public. Plans for a prototype have been delayed somewhat, while we assess ways that the project might interact with the National Science Foundation's CHRONOS project (described in a paper by Wardlaw in this volume).

PHASE TWO

Phase Two focuses on development of standards and guidelines needed to assist the USGS and state geological surveys in efficiently producing digital geologic maps, in a more standardized and common format. Our profession encourages innovation and individual pursuit of science, and so the question may be posed – why do we need these standards? Clearly, standards should not impede science but instead should help us efficiently communicate our science to the public. The need for communication was perhaps best articulated by former USGS Director John Wesley Powell, while planning for the new Geologic Atlas of the United States:

> *"... the maps are designed not so much for the specialist as for the people, who justly look to the official geologist for a classification, nomenclature, and system of convention so simple and expressive as to render his work immediately [understandable]..." (Powell, 1888).*

At that time, and throughout the early 20th century, Powell and others guided the USGS and the Nation's geoscientists toward a set of robust, practical standards for classifying geologic units and materials and representing them on maps. Those standards endured and evolved, and continue as basic guidelines for geologic mapping. Although today we commonly record in the field and laboratory far more complex information than during Powell's era, the necessity to provide it to the public in a standardized format remains unchanged. Newly evolving data formats and display techniques made feasible by computerization challenge us to revisit Powell's vision, and to develop standards and guidelines appropriate to today's technology and science.

In mid-1996, the NGMDB project and the AASG convened a meeting to identify the types of standards and guidelines that would improve the quality and utility of digital maps produced by the nation's geological surveys. From that meeting, Standards Working Groups were formed to address: 1) standard symbolization on geologic maps; 2) standard procedures for creating digital maps; 3) guidelines for publishing digital geologic maps; 4) documentation of methods and information via formal metadata; and 5) standard data structures and science terminology for geologic databases. The working group results will help provide a set of national standards to support public use of standard, seamless geologic map information for the entire country. In essence, Powell's pragmatic vision for the Geologic Atlas of the U.S. has been applied a century later to the National Geologic Map Database.

The tasks assigned to these Standards Working Groups are interrelated, as shown in figure 9 – when in the field, a geologist makes observations and (often, provisionally) draws geologic features on a base map; at that time, the accuracy with which these features are located on the map can be estimated. Further, the information may be recorded digitally in the field; if so, it can be structured similar to, or compatible with, the map database's structure (the "data model" in this figure). Returning to the office, the geologist commonly organizes and interprets field observations and prepares for map production – descriptions may be standardized according to an agency or project-level terminology or "science language," the map data may be structured according to the standard data model implemented by the agency, and procedures may be documented with metadata both in the office and when gathering data in the field. The descriptive information then is combined with the feature location information in a GIS, and digital cartography is applied to create a map that is published according to agency policies. Finally, the map is released to the public and accessed through various mechanisms including the NGMDB.

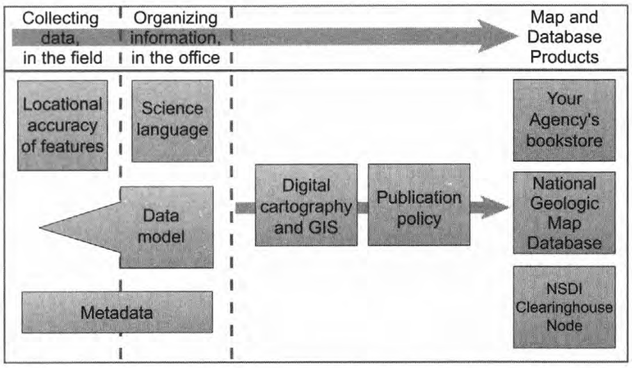


Figure 9. Diagram showing how the standards and guidelines under development by the NGMDB and related groups relate to the process of making a map.

As described below, since 1996 these Working Groups and their successor organizations have made significant progress toward developing some of the necessary standards and guidelines. General information about the Working Groups and details of their activities are available at <http://ncgmp.usgs.gov/ngmdbproject/standards/>. Working Group members are listed in Appendix A.

Internationally, the NGMDB participates in venues that help to develop and refine the U.S. standards. These venues also bring our work to the international community, thereby promoting greater standardization with other countries. Examples include:

1. participation in "DIMAS", the map standards committee of the Commission for the Geological Map of the World (see article in this volume, and http://www.geology.cz/host/dimas.htm), and
2. development of a map database and standards Clearinghouse (http://ncgmp.usgs.gov/intdb/) that is endorsed by the International Union of Geological Sciences' Commission for the Management and Application of Geoscience Information ("CGI", http://www.iugs.org/iugs/science/sci-cnfo.htm) and the

International Association for Mathematical Geology (http://www.iamg.org/).

Geologic Map Symbolization

A draft standard for geologic map line and point symbology and map patterns and colors, published in a USGS Open-File Report in 1995, was reviewed in 1996 by the AASG, USGS, and Federal Geographic Data Committee (FGDC). It was revised by the NGMDB project team and members of the USGS Western Region Publications Group, and in late 1997 was circulated for internal review. The revised draft then was prepared as a proposed federal standard, for consideration by the FGDC. The draft was, in late 1999 through early 2000, considered and approved for public review by the FGDC and its Geologic Data Subcommittee. The document was released for public comment within the period May 19 through September 15, 2000 (see http://ncgmp.usgs.gov/fgdc_gds/mapsymb/ for the document and for information about the review process). This draft standard is described in some detail in Soller and Lindquist (2000). Based on public review comments, in 2002 a new section was added to the draft standard to address uncertainty in locational accuracy of map features. This section was presented for comment (Soller and others, 2002) and revised accordingly. With assistance from a Standing Committee to oversee resolution of review comments and long-term maintenance of the standard, the document is being prepared for submittal to FGDC, for final discussion and adoption as a Federal standard.

Digital Mapping

The Data Capture Working Group has coordinated seven annual "Digital Mapping Techniques" (DMT) workshops for state, federal, and Canadian geologists, cartographers, managers, and industry partners. These informal meetings serve as a forum for discussion and information-sharing, and have been quite successful. They have significantly helped the geoscience community converge on more standardized approaches for digital mapping and GIS analysis, and thus agencies have adopted new, more efficient techniques for digital map preparation, analysis, and production. In support of DMT workshops, an email listserver is maintained to facilitate the exchange of specific technical information.

The most recent DMT workshop, held in Millersville, Pennsylvania, and hosted by the Pennsylvania Geological Survey, was attended by 90 representatives of 36 state, federal, and Canadian agencies and private companies. Workshop proceedings are published (see Appendix B and http://ncgmp.usgs.gov/ngmdbproject/standards/datacapt/). Published copies of the proceedings may be obtained from David Soller or Thomas Berg.

Map Publication Requirements

Through the USGS Geologic Division Information Council, the NGMDB led development of the USGS policy "Publication Requirements for Digital Map Products" (enacted May 24, 1999; see link under Map Publication Guidelines, at http://ncgmp.usgs.gov/ngmdbproject/standards/). A less USGS-specific version of this document was developed by the Data Information Exchange Working Group and presented for technical review at a special session of the Digital Mapping Techniques '99 workshop (Soller and others, 1999a). The revised document (entitled "Proposed Guidelines for Inclusion of Digital Map Products in the National Geologic Map Database") was reviewed by the AASG Digital Geologic Mapping Committee. In 2002, it was unanimously approved via an AASG resolution, and has been incorporated as a guideline for digital map product deliverables to the STATEMAP component of the National Cooperative Geologic Mapping Program (see link under Map Publication Guidelines, at http://ncgmp.usgs.gov/ngmdbproject/standards/).

Among the geological surveys there are many approaches to determining authorship credit and citation format for geologic maps, digital geologic maps, and associated databases. It is prudent for agencies to adopt policies that preserve the relationship of the geologist-authors to their product, the map image, and to identify the appropriate authorship (if any) and/or credit for persons responsible for creating the database files. A summary of this issue and a proposed guideline was discussed at the Digital Mapping Techniques workshop in 2001 (Berquist and Soller, 2001).

Metadata

The Metadata Working Group developed its final report in 1998. The report provides guidance on the creation and management of well-structured formal metadata for digital maps (see http:// ncgmp.usgs.gov/ngmdbproject/standards/metadata/metaWG.html). The report contains links to metadata-creation tools and general discussions of metadata concepts (see, for example, the metadata-creation tools, "Metadata in Plain Language," and other helpful information at <http://geology.usgs.gov/tools/metadata/>.

Geologic Map Data Model

In early 1999, the Data Model Working Group had concluded its work with release of a draft version of a data model (Johnson and others, 1998). The Group then was succeeded by the North American Data Model Steering Committee (NADMSC, http://geology.usgs.gov/dm/). State and USGS collaborators on the NGMDB continue to participate in this activity, helping to develop, refine, and test the North American Geologic Map Data Model ("NADM") and the standard science language that must accompany it. This work recently has produced a significant accomplishment, the NADM Conceptual Data Model. This model is available for perusal and comment, at <http://geology.usgs.gov/dm/steering/teams/design/NADM-C1.0/NADMC1_0.pdf>. Information about other Committee activities is provided in two papers in this volume: 1) the development of a XML-based interchange format; and 2) the development of standard science language to describe the lithology of earth materials.

To provide templates for building GIS data, ESRI is designing ArcGIS data models for many industries and applications (see http://esri.com/software/arcgisdatamodels/index.html). Through discussions that involved the NGMDB, ESRI plans to structure the ArcGIS data model at least in part on concepts in the NADM Data Model.

PHASE THREE

Over the past few decades, significant advances in computer technology have begun to permit complex spatial information (especially vector-based) to be stored, managed, and analyzed for use by a growing number of geoscientists. At the beginning of the NGMDB project, we judged that computer-based mapping was not a sufficiently mature discipline to permit us to develop an online, vector-based map database. In particular, technology for display and query of complex spatial information on the Web was in its infancy, and hence was not seriously considered by the NGMDB project as a viable means to deliver information to the general public. However, there now exists sufficient digital geologic map data; sufficient convergence on standard data formats, data models, GIS and digital cartographic practices and field data capture techniques; and sufficient technological advances in Internet delivery of spatial information to warrant a research effort for a prototype, online vector-based map database.

Before beginning to design this database, project personnel held numerous discussions with geoscientists and the general public to gauge interest in an online database and to define its scope. Based on these discussions, it was clear that this database should be:

1. built from edge-matched geologic maps at various scales;
2. managed and accessed as a coherent body of map information, not just as a set of discrete map products;
3. updated by mappers and/or a committee, "on the fly" when new information becomes available - it should be a "living" database;
4. standardized, adhering to a standard data model with standard scientific terminology; and
5. available to users via Internet browsers and common GIS tools (such as ArcExplorer).

This database will integrate with other databases developed under the NGMDB project. For example, a user accessing the online, vector-based map database might identify a map unit of interest, and then want to purchase or download the original published map product, or inquire about fossils found within that unit, or learn about the history of the geologic unit. Also, a user might access the Map Catalog and identify a map of interest, and then be linked to the online map database in order to browse and query it.

Prototyping

The NGMDB project has begun a series of prototypes, to advance our understanding of the technical and management challenges to developing the operational system; an introduction is given in Soller and others (2000). In 1999, we outlined some basic requirements for the prototype and tested them using map data for the greater Yellowstone area of Wyoming and Montana (Wahl and others, 2000). The second prototype (Soller and others, 2001) was conducted in cooperation with the Kentucky Geological Survey. In this prototype, we demonstrated in a commercial database system (GE-Smallworld) how the geologic database could be analyzed over the Web in concert with local datasets. The data model for the second prototype is described in Soller and others (2002), and was a significant contributor to the design of the new NADM Conceptual Data Model noted above.

Before proceeding further with plans for the publicly-accessible map database, we need to define a set of standardized terminology for the properties of earth materials (the science language). This science language must be sufficiently robust to accommodate terminology generated through today's field mapping, and terminology found in map unit descriptions on older and on smaller-scale maps, where descriptions tend to be highly generalized. Also, we need to collect enough standardized geologic map data to justify the cost of developing the database. Therefore, in our third prototype we will create map data with a standardized data model and science language, using available mapping in disparate field areas (central Arizona, northern Virginia, Kentucky, southern California, and the Greater Yellowstone Area; see fig. 10). To achieve this, we are writing data-entry software tools supported by science language derived from the NADMSC.

What is a data model, and how does it apply to geologic maps?

A data model provides organization to the descriptive and spatial information that constitute a geologic map.

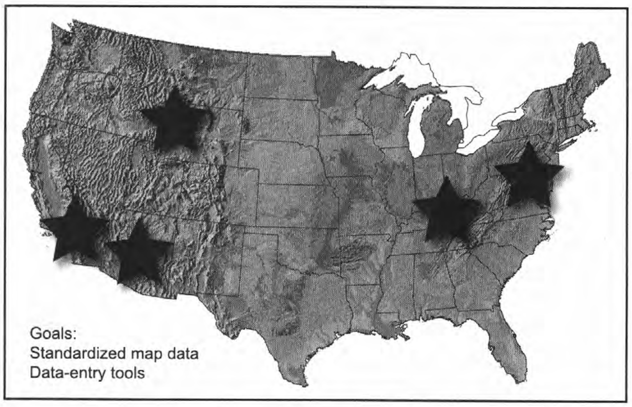


Figure 10. The goals of the current prototype are to: 1) create map data that has a standardized data model and science language, beginning with some national-scale maps and available mapping in disparate field areas shown above, and 2) create data-entry tools that are flexible and readily modified, enabling geologists to enter detailed, more standardized descriptive information.

The relations between a data model, science language, and the geologic map require some explanation. A data model may be highly conceptual, or it may describe the data structure for managing information within a specific hardware/software platform. In either case, it is a central construct because it addresses the database design for geologic maps in GIS format. In figure 11, the data model is simplified to four locations, or "bins", where information can be stored, with each bin containing many database tables and fields:

1. Occurrence – this bin contains the spatial geometry for each geologic feature in a map database. For example, the map unit identifier and the coordinates that define the outline of a map unit are included here.
2. Descriptor – this bin contains the wealth of descriptive information for each feature that occurs in the map database. This can include the full map unit description and simple attributes such as dominant lithology, color, and the nature of bedding.
3. Concept – this bin contains essential reference standards, such as geologic time scale(s) and science language. It also contains concepts and definitions essential for querying the database (for example, the concept that a rock can "intrude" another rock).
4. Symbol – this bin includes cartographic entities for symbolizing the map on-screen and in print form.

Will the U.S. have a single standard data model and science language?

The NGMDB online map database is envisioned as a distributed system that will provide seamless access to, and display of, map data served by many agencies. If all agencies used the same science language and exactly the same data model, and if it were implemented on the same hardware and software platform, a functional system would be relatively easy to build. That, however, is not a realistic scenario. Each agency has a unique history, set of objectives, and budget that will dictate the nature of their map database. (It should be noted that not all geological surveys in the U.S. can now afford to build such a system.) A more realistic approach is to assume a heterogenous computing environment, and to build software that can translate data structure and science language from one agency's system to another (fig. 12). This translation mechanism ensures "interoperability" between systems, and is the most realistic approach for the NGMDB.

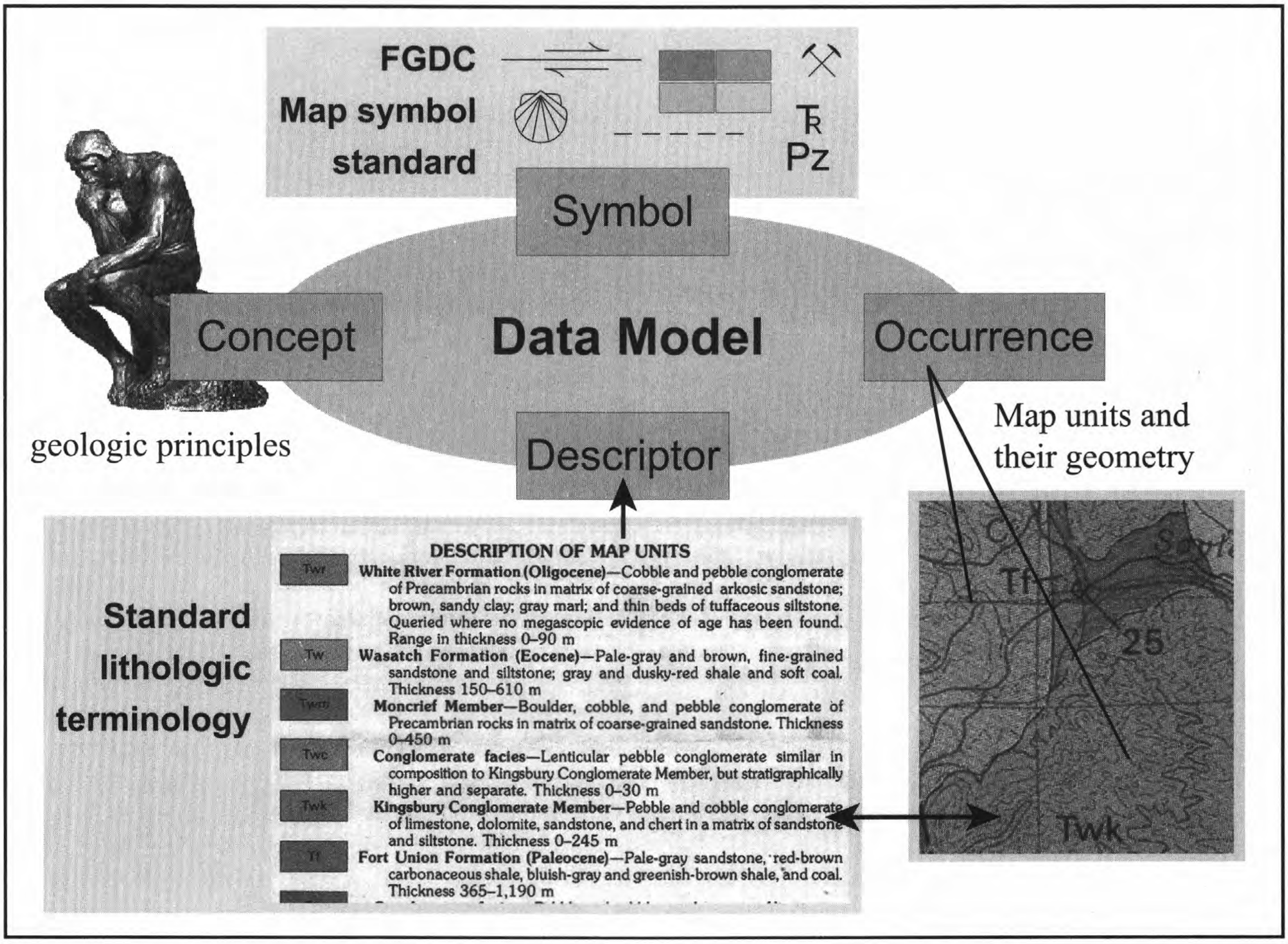
To facilitate interoperability among systems, the NGMDB will define and maintain a set of reference standards (for data model, science language, time scale) based in part on those produced by the NADMSC. Interoperability software that enables disparate systems to appear to the user as a single system is being evaluated by groups including the NADMSC, NGMDB, and the National Science Foundation-funded GEON project. We anticipate collaborative research, especially with GEON, on XML-based "wrapper/mediator" technology to address these needs for the NGMDB. Through this technology, agencies should be able to correlate their unique data structure and scientific terminology to the reference standard, and the translator (presumably XML-based) will enable us to display the information to the user in a single view.

Extending the data model to include three-dimensional (3-D) map information

The data model was designed for the typical geologic map, which provides a two-dimensional representation of the geologic framework. On most geologic maps, this framework is expressed generally, in cross-sections and map unit descriptions. The data model can accommodate more detailed and location-specific 3-D information, although it has not yet been applied in this fashion.

Three-dimensional geologic map information can be represented by various methods. The most traditional approach is vector-based stack-unit mapping, where a vertical stack of surface and subsurface geologic units are combined into a two-dimensional (2-D) map unit (fig. 13a). The stack-unit characterizes the vertical variations of physical properties in each 3-D map unit. These maps are readily managed in the data model, like a traditional geologic map (fig. 11).

Map unit descriptions, whether on traditional 2-D geologic maps or vector-based stack-unit maps, apply to the entire unit. As a consequence, if a map unit's texture is described as "generally sandy, although fining to the east," the unit cannot be readily subdivided into areas that are sandy and those that are finer. This can be a limita-



DESCRIPTION OF MAP UNITS

White River Formation (Oligocene)—Cobble and pebble conglomerate of Precambrian rocks in matrix of coarse-grained arkosic sandstone; brown, sandy clay; gray marl; and thin beds of tuffaceous siltstone. Queried where no megascopic evidence of age has been found. Range in thickness 0–90 m

Wasatch Formation (Eocene)—Pale-gray and brown, fine-grained sandstone and siltstone; gray and dusky-red shale and soft coal. Thickness 150–610 m

Moncrief Member—Boulder, cobble, and pebble conglomerate of Precambrian rocks in matrix of coarse-grained sandstone. Thickness 0–450 m

Conglomerate facies—Lenticular pebble conglomerate similar in composition to Kingsbury Conglomerate Member, but stratigraphically higher and separate. Thickness 0–30 m

Kingsbury Conglomerate Member—Pebble and cobble conglomerate of limestone, dolomite, sandstone, and chert in a matrix of sandstone and siltstone. Thickness 0–245 m

Fort Union Formation (Paleocene)—Pale-gray sandstone, red-brown carbonaceous shale, bluish-gray and greenish-brown shale, and coal. Thickness 365–1,190 m

Figure 11. Simplified representation of the data model and its application to a typical, 2-D geologic map. The presence of a geologic unit on the map, referred to in the data model as an "occurrence" of that map unit, is described by: 1) its bounding contacts and faults, whose coordinates are stored as the unit's "geometry"; and 2) its physical properties, which are stored as the unit's "descriptors."

tion to users, especially when using the map for detailed studies. In contrast to vector-based stack-unit maps, voxel maps show every part of a geologic unit as a unique point known as a volume-pixel or voxel. Each voxel can have a unique set of attributes, therefore lateral and vertical variations in texture within the geologic unit can be described in great detail. Such information is difficult to collect at depth, and so in studies where this type of representation is needed, voxel attributes tend to be computed from a few point measurements within the geologic unit.

A third approach to 3-D mapping, raster-based stacked surfaces, offers a useful compromise between vector-based stack-unit and voxel-based mapping. In this approach, a set of 2-D elevation maps shows, in raster format, the surface of each buried geologic unit. These surfaces are in many cases rasterized from conventional, vector-based maps. Unlike the vector-based stack-unit map, they provide the opportunity to model the surface elevation and thickness of each unit, and to assign unique physical properties to each location on the unit's surface. Although not as detailed as a voxel representation, this approach requires less information and fewer assumptions about the 3-D variation of properties within the unit, and can more readily be created using conventional GIS software such as ArcGIS. Lateral variations in a physical property such as texture can be recorded; this is informative for units such as alluvium, which may have distinct subenvironments with different characteristics (for example, coarser material in the main channel, and finer material in overbank areas and tributaries).

Raster-based stacked surfaces (and, by extension, voxel-based maps) can be represented in the data model, as shown in figure 13b. This raster-specific information can significantly improve the value of geologic data when applied to, for example, groundwater modeling. The 3-D geometry of the glacial aquifer shown in figure 13b was provided to a private groundwater consortium in order to develop a regional groundwater flow model. The aquifer is composed of coarse sand and gravel in the main channel but is finer-grained in the tributaries because sedi-

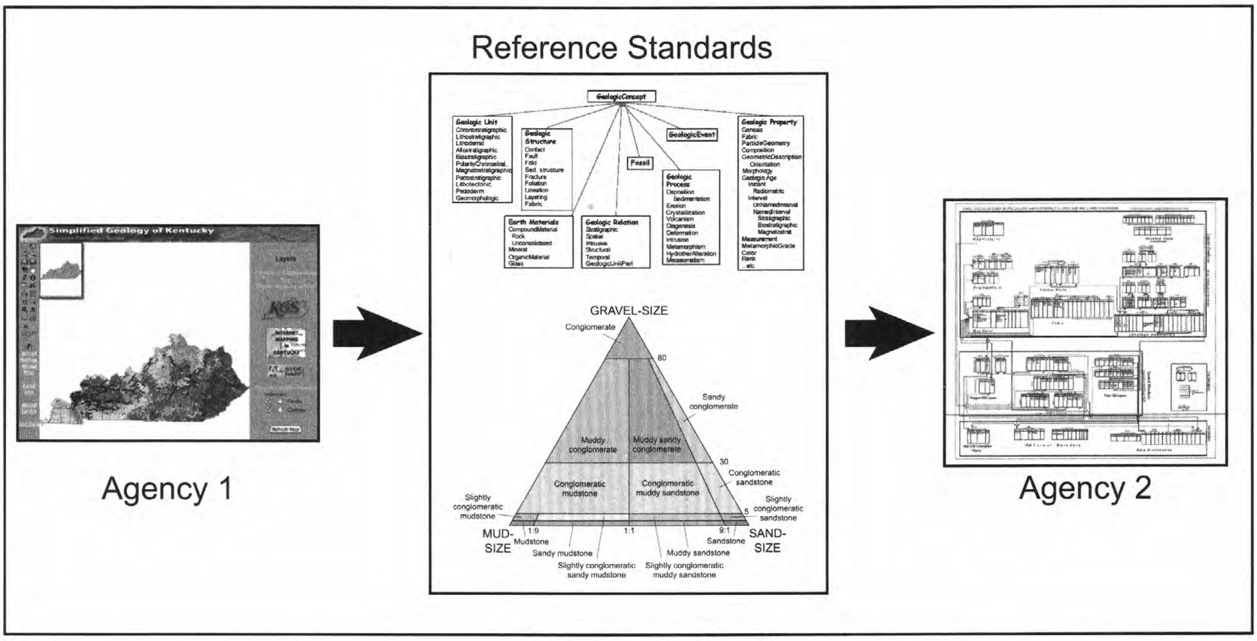


Figure 12. A single, monolithic system design shared by all agencies is unlikely. Rather, interoperability among the many agency databases linked together by the NGMDB database is the most logical design philosophy. In this diagram, we envision that map data from one agency (the Kentucky Geological Survey, http://www.uky.edu/KGS/) will be translated into reference standards (the data model and science language standards adopted by the NGMDB) and translated out to the criteria required by another agency (the Idaho Geological Survey's Geologic Map Data Model, http://www.idahogeology.org/Lab/datamodel.htm). This approach also could permit the NGMDB to coordinate the translation and display of multiple agency databases. In this diagram, the reference standards are represented by a schematic of the draft NGMDB data model (discussed in another paper in this volume) and an example of science language from Folk (1954, fig. 1a) showing a rock classification based on mud-sand-gravel content.

ment dammed the margins of the main channel, causing lakes to form in tributaries. When the 3-D information was provided to the consortium, the authors did not have sufficient data to assign to the units any lateral variations in texture. As a result, the groundwater modelers had to assume a homogenous aquifer. Raster surfaces that showed lateral variations in sediment texture would have enabled the modelers to consider the heterogeneity that was known to exist within that aquifer.

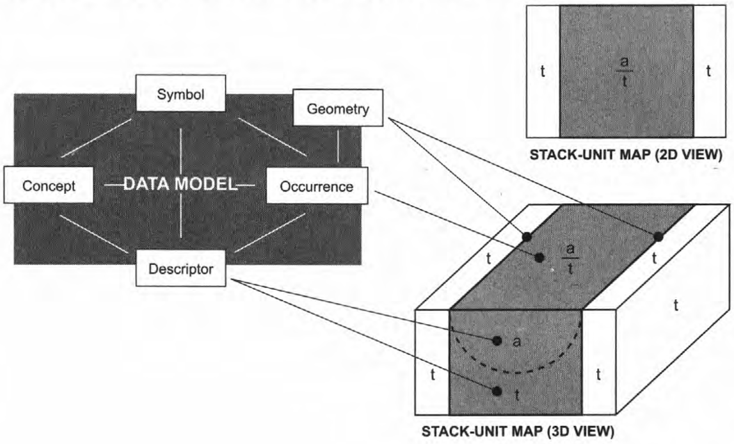
National and regional map coverage

The online map database will be more useful if it includes some geologic map coverage for the entire nation. To that end, the NGMDB has supported compilation and GIS development of several regional maps (fig. 14). Most significant is the digital version of the "Geologic Map of North America". This map is the final product of the Geological Society of America's (GSA) Decade of North American Geology project. The NGMDB has provided funding and expertise for development of the digital files that will be used to print the map, in order to engage GSA in a plan to develop a database for the map. When compilation and review of the map has been completed, hopefully within the next year, we will propose a database design and begin to populate the digital files made available from cartographic production of the map. This work will be conducted in collaboration with GSA and interested national geological surveys. The other maps shown in figure 14 are published or in press, and we intend to process these for inclusion in the online map database.

ACKNOWLEDGMENTS

The authors thank the members of the NGMDB project staff and collaborators for their enthusiastic and expert support, without which the project would not be possible. In particular, we thank: Nancy Stamm (USGS, Reston, VA; Geolex database, and general support and guidance to the project), Ed Pfeifer, Alex Acosta, Dennis McMacken, Jana Ruhlman, and Michael Gishey (USGS, Flagstaff, AZ; Website and database management), Chuck Mayfield and Nancy Blair (USGS Library; Map Catalog content), Bruce Wardlaw (USGS, Reston, VA; Paleontology Database), Robert Wardwell (USGS, Vancouver, WA; Image Library) and Kevin Laurent and Jeremy Skog (USGS, Reston, VA; Image Library), Steve Richard (Arizona Geological Survey, Tucson, AZ; data model and

A. VERTICAL VARIATIONS OF PHYSICAL PROPERTIES IN A MAP UNIT



B. LATERAL VARIATIONS OF PHYSICAL PROPERTIES IN A MAP UNIT

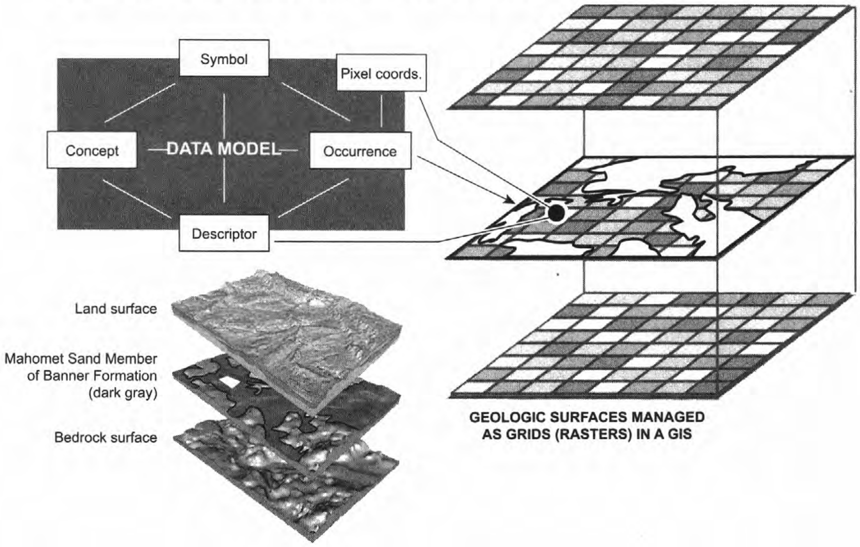


Figure 13. Approaches for representing three-dimensional map information, and for managing it in the data model.

A. Vector-based stack-unit maps depict the vertical succession of geologic units to a specified depth (here, the base of the block diagram). This mapping approach characterizes the vertical variations of physical properties in each 3-D map unit. In this example, an alluvial deposit (unit "a") overlies glacial till (unit "t"), and the stack-unit labeled "a/t" indicates that relationship, whereas the unit "t" indicates that glacial till extends down to the specified depth. In a manner similar to that shown in figure 11, the stack-unit's occurrence (the map unit's outcrop), geometry (the map unit's boundaries), and descriptors (the physical properties of the geologic units included in the stack-unit) are managed as they are for a typical 2-D geologic map.

B. Raster-based stacked surfaces depict the surface of each buried geologic unit, and can accommodate data on lateral variations of physical properties. In this example from Soller and others (1999), the upper surface of each buried geologic unit was represented in raster format as an ArcInfo Grid file. The middle grid is the uppermost surface of an economically important aquifer, the Mahomet Sand, which fills a pre- and inter-glacial valley carved into the bedrock surface. Each geologic unit in raster format can be managed in the data model, in a manner not dissimilar from that shown for the stack-unit map. The Mahomet Sand is continuous in this area, and represents one occurrence of this unit in the data model. Each raster, or pixel, on the Mahomet Sand surface has a set of map coordinates that are recorded in a GIS (in the data model bin that is labeled "Pixel coordinates", which is the raster corollary of the "Geometry" bin for vector map data). Each pixel can have a unique set of descriptive information, such as surface elevation, unit thickness, lithology, transmissivity, etc.).

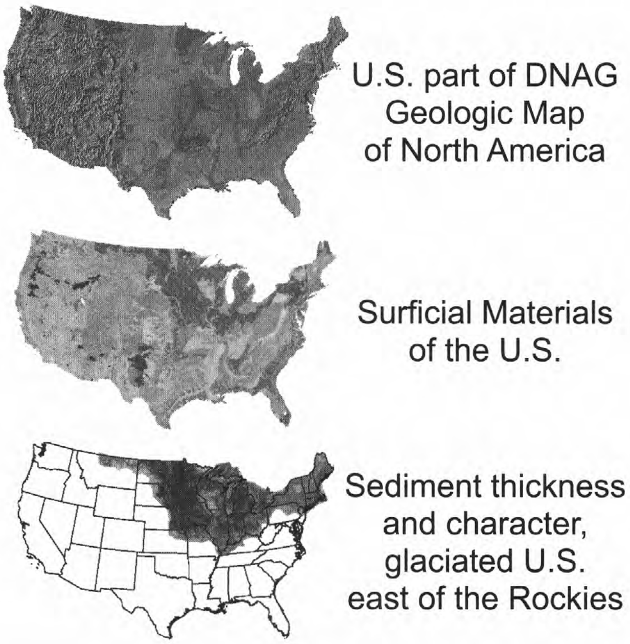


Figure 14. Regional maps whose compilation and/or GIS development is supported by the NGMDB. The uppermost map, the Geologic Map of North America, is discussed in the text. The center map is in press (Soller and Reheis, in press) and must be converted to a database. The database for the lower map is published (Soller and Packard, 1998) and will be adapted to the emerging NGMDB standards.

science language), Jonathan Matti (USGS, Tucson, AZ; data model and science language), and Jordan Hastings (USGS, Santa Barbara, CA; data model).

We also thank the many committee members who provided technical guidance and standards (App. 1).

REFERENCES

Berquist, C.R., Jr., and Soller, D.R., 2001, Map authorship and citation guidelines: Summary of a discussion session, *in* Soller, D.R., ed., Digital Mapping Techniques '01—Workshop proceedings: U.S. Geological Survey Open-file Report 01-223, p. 159-164, <http://pubs.usgs.gov/of/2001/of01-223/berquist.html>.

Folk, R.L., 1954, The distinction between grain size and mineral composition in sedimentary rock nomenclature: Journal of Geology, v. 62, p. 344-359.

Johnson, B.R., Brodaric, B., Raines, G.L., Hastings, J., and Wahl, R., 1998, Digital geologic map data model v. 4.3: AASG/USGS Data Model Working Group Report, <http://geology.usgs.gov/dm/>.

Mac Lachlan, M.E., and others, compilers, 1996, Stratigraphic nomenclature databases for the United States, its possessions, and territories: U.S. Geological Survey Digital Data Series DDS-6, Release 3, one CD-ROM.

Powell, J.W., 1888, Methods of geologic cartography in use by the United States Geological Survey, *in* Congrès Géologique International, Compte Rendu de la 3me Session, Berlin, 1885: A.W. Schade's Buchdruckerei, Berlin, p. 221–240.

Soller, D.R., Berg, T.M., and Wahl, Ron, 2000, Developing the National Geologic Map Database, phase 3—An online, "living" database of map information, *in* Soller, D.R., ed., Digital Mapping Techniques '00—Workshop Proceedings: U.S. Geological Survey Open-File Report 00-325, p. 49-52, <http://pubs.usgs.gov/openfile/of00-325/soller4.html>.

Soller, D.R., Brodaric, Boyan, Hastings, J.T., Wahl, Ron, and Weisenfluh, G.A., 2002, The central Kentucky prototype: An object-oriented geologic map data model for the National Geologic Map Database: U.S. Geological Survey Open-file Report 02-202, 38 p., <http://pubs.usgs.gov/of/2002/of02-202/>.

Soller, D.R., Duncan, Ian, Ellis, Gene, Giglierano, Jim, and Hess, Ron, 1999, Proposed guidelines for inclusion of digital map products in the National Geologic Map Database, *in* Soller, D.R., ed., Digital Mapping Techniques '99—Workshop Proceedings: U.S. Geological Survey Open-File Report 99-386, p. 35-38, <http://pubs.usgs.gov/openfile/of99-386/soller2.html>.

Soller, D.R., and Lindquist, Taryn, 2000, Development and public review of the draft "Digital cartographic standard for geologic map symbolization", *in* Soller, D.R., ed., Digital Mapping Techniques '00—Workshop Proceedings: U.S. Geological Survey Open-File Report 00-325, p. 43-47, <http://pubs.usgs.gov/openfile/of00-325/soller3.html>.

Soller, D.R., Lindquist, T.A., and Matti, J.C., 2002, Field description of the scientific and locational accuracy of geologic features (a part of the draft FGDC geologic map symbolization standard), *in* Soller, D.R., ed., Digital Mapping Techniques '02—Workshop Proceedings: U.S. Geological Survey Open-File Report 02-370, p. 33-40, <http://pubs.usgs.gov/of/2002/of02-370/soller1.html>.

Soller, D.R., and Packard, P.H., 1998, Digital representation of a map showing the thickness and character of Quaternary sediments in the glaciated United States east of the Rocky Mountains: U.S. Geological Survey Digital Data Series DDS-38, one CD-ROM, <http://pubs.usgs.gov/dds/dds38/>.

Soller, D.R., Price, S.D., Kempton, J.P., and Berg, R.C., 1999, Three-dimensional geologic maps of Quaternary sediments in east-central Illinois: U.S. Geological Survey Geologic Investigations Series Map I-2669, <http://pubs.usgs.gov/imap/i-2669/>.

Soller, D.R., and Reheis, M.C., compilers, in press, Surficial materials in the conterminous United States: U.S. Geological Survey Open-file Report 03-275, scale 1:5,000,000.

Soller, D.R., Wahl, Ron, Weisenfluh, Jerry, Brodaric, Boyan, Hastings, Jordan, Laudati, Robert, and Fredericks, Roger, 2001, Progress report on the National Geologic Map Database, phase 3: An online database of map information, *in* Soller, D.R., ed., Digital Mapping Techniques '01—Workshop proceedings: U.S. Geological Survey Open-file Report 01-223, p. 71-78, <http://pubs.usgs.gov/of/2001/of01-223/soller2.html>.

Stamm, N.R., Wardlaw, B.R., and Soller, D.R., 2000, GEOLEX – The National Geologic Map Database's geologic names lexicon, *in* Soller, D.R,. ed., Digital Mapping Techniques '00—Workshop Proceedings: U.S. Geological Survey

Open-File Report 00-325, p. 31-42, <http://pubs.usgs.gov/of/2000/of00-325/stamm.html >.

Wahl, R.R., Soller, D.R., and Yeldell, Steven, 2000, Prototype Implementation of the NADMSC Draft Standard Data Model, Greater Yellowstone Area, *in* Soller, D.R., ed., Digital Mapping Techniques '00—Workshop Proceedings: U.S. Geological Survey Open-File Report 00-325, p. 57-63, <http://pubs.usgs.gov/openfile/of00-325/wahl.html>.

Wardlaw, B.R., Stamm, N.R., and Soller, D.R., 2001, The U.S. Geological Survey National Paleontological Database, *in* Soller, D.R., editor, Digital Mapping Techniques '01—Workshop proceedings: U.S. Geological Survey Open-file Report 01-223, p. 59-70, <http://pubs.usgs.gov/of/2001/of01-223/wardlaw.html>.

APPENDIX A

Principal committees and people collaborating with the National Geologic Map Database project.

Digital Geologic Mapping Committee of the Association of American State Geologists:
Tom Berg (Ohio Geological Survey and Committee Chair)
Rick Allis (Utah Geological Survey)
Lee Allison (Kansas Geological Survey)
Larry Becker (Vermont Geological Survey)
Rick Berquist (Virginia Division of Mineral Resources)
Jim Cobb (Kentucky Geological Survey)
Ian Duncan (Virginia Division of Mineral Resources)
Rich Lively (Minnesota Geological Survey)
Jay Parrish (Pennsylvania Geological Survey)
Bill Shilts (Illinois State Geological Survey)
Nick Tew (Alabama Geological Survey)
Harvey Thorleifson (Minnesota Geological Survey)

Technical Advisory Committee:
Boyan Brodaric (Geological Survey of Canada)
David Collins (Kansas Geological Survey)
Larry Freeman (Alaska Division of Geological & Geophysical Surveys)
Jordan Hastings (University of California, Santa Barbara)
Dan Nelson (Illinois State Geological Survey)
Stephen Richard (Arizona Geological Survey)
Jerry Weisenfluh (Kentucky Geological Survey)

Map Symbol Standards Committee:
Dave Soller (U.S. Geological Survey and Committee Coordinator)
Tom Berg (State Geologist, Ohio Geological Survey)
Bob Hatcher (University of Tennessee, Knoxville)
Mark Jirsa (Minnesota Geological Survey)
Taryn Lindquist (U.S. Geological Survey)
Jon Matti (U.S. Geological Survey)
Jay Parrish (State Geologist, Pennsylvania Geological Survey)
Jack Reed (U.S. Geological Survey)
Steve Reynolds (Arizona State University)
Byron Stone (U.S. Geological Survey)

AASG/USGS Data Capture Working Group:
Dave Soller (U.S. Geological Survey and Group Chair)
Warren Anderson (Kentucky Geological Survey)
Rick Berquist (Virginia Division of Mineral Resources)
Elizabeth Campbell (Virginia Division of Mineral Resources)
Rob Krumm (Illinois State Geological Survey)
Scott McCulloch (West Virginia Geological and Economic Survey)
Gina Ross (Kansas Geological Survey)
George Saucedo (California Geological Survey)
Barb Stiff (Illinois State Geological Survey)
Tom Whitfield (Pennsylvania Geological Survey)

DMT Listserve:
Maintained by Doug Behm, University of Alabama

AASG/USGS Metadata Working Group:
Peter Schweitzer (U.S. Geological Survey and Group Chair)
Dan Nelson (Illinois State Geological Survey)
Greg Hermann (New Jersey Geological Survey)
Kate Barrett (Wisconsin Geological and Natural History Survey)
Ron Wahl (U.S. Geological Survey)

AASG/USGS Data Information Exchange Working Group:
Dave Soller (U.S. Geological Survey and Group Chair)
Ron Hess (Nevada Bureau of Mines and Geology)
Ian Duncan (Virginia Division of Mineral Resources)
Gene Ellis (U.S. Geological Survey)
Jim Giglierano (Iowa Geological Survey)

AASG/USGS Data Model Working Group:
Gary Raines (U.S. Geological Survey and Group Chair)
Boyan Brodaric (Geological Survey of Canada)
Jim Cobb (Kentucky Geological Survey)
Ralph Haugerud (U.S. Geological Survey)
Greg Hermann (New Jersey Geological Survey)
Bruce Johnson (U.S. Geological Survey)
Jon Matti (U.S. Geological Survey)
Jim McDonald (Ohio Geological Survey)
Don McKay (Illinois State Geological Survey)
Steve Schilling (U.S. Geological Survey)
Randy Schumann (U.S. Geological Survey)
Bill Shilts (Illinois State Geological Survey)
Ron Wahl (U.S. Geological Survey)

North American Data Model Steering Committee:
Dave Soller (U.S. Geological Survey and Committee Coordinator)
Tom Berg (Ohio Geological Survey)
Boyan Brodaric (Geological Survey of Canada and Chair of the Data Model Design Technical Team)
Bruce Johnson (U.S. Geological Survey and Chair of the Data Interchange Technical Team)
Murray Journeay (Geological Survey of Canada)
Rob Krumm (Illinois State Geological Survey)
Jonathan Matti (U.S. Geological Survey and Chair of the Science Language Technical Team)
Scott McColloch (West Virginia Geological and Economic Survey)
Steve Richard (Arizona Geological Survey)
Peter Schweitzer (U.S. Geological Survey)
Loudon Stanford (Idaho Geological Survey)
Jerry Weisenfluh (Kentucky Geological Survey)

NGDMB contact-persons in each State geological survey:
These people help the NGMDB with the Geoscience Map Catalog, GEOLEX, the Geologic Map Image Library, and the Mapping in Progress Database. Please see <http://ncgmp.usgs.gov/ngmdbproject/statecontacts.html> for this list.

APPENDIX B

List of progress reports on the National Geologic Map Database,
and Proceedings of the Digital Mapping Techniques workshops.

Soller, D.R., editor, 2002, Digital Mapping Techniques '02—Workshop Proceedings: U.S. Geological Survey Open-File Report 02-370, 214 p., <http://pubs.usgs.gov/of/2002/of02-370/>.

Soller, D.R., editor, 2001, Digital Mapping Techniques '01—Workshop Proceedings: U.S. Geological Survey Open-File Report 01-223, 248 p., <http://pubs.usgs.gov/of/2001/of01-223/>.

Soller, D.R., editor, 2000, Digital Mapping Techniques '00—Workshop proceedings: U.S. Geological Survey Open-file Report 00-325, 209 p., <http://pubs.usgs.gov/openfile/of00-325/>.

Soller, D.R., editor, 1999, Digital Mapping Techniques '99—Workshop proceedings: U.S. Geological Survey Open-file Report 99-386, 216 p., <http://pubs.usgs.gov/openfile/of99-386/>.

Soller, D.R., editor, 1998, Digital Mapping Techniques '98—Workshop Proceedings: U.S. Geological Survey Open-File Report 98-487, 134 p., <http://pubs.usgs.gov/openfile/of98-487/>.

Soller, D.R., editor, 1997, Proceedings of a workshop on digital mapping techniques: Methods for geologic map data capture, management, and publication: U.S. Geological Survey Open-File Report 97-269, 120 p., <http://ncgmp.usgs.gov/pubs/of97-269/>.

Soller, D.R., and Berg, T.M., 2002, The National Geologic Map Database: A progress report, *in* Soller, D.R., editor, Digital Mapping Techniques '02—Workshop proceedings: U.S. Geological Survey Open-file Report 02-370, p. 75-83, <http://pubs.usgs.gov/of/2002/of02-370/soller2.html>.

Soller, D.R., and Berg, T.M., 2001, The National Geologic Map Database—A progress report, *in* Soller, D.R., editor, Digital Mapping Techniques '01—Workshop proceedings: U.S. Geological Survey Open-file Report 01-223, p. 51-57, <http://pubs.usgs.gov/of/2001/of01-223/soller1.html>.

Soller, D.R., and Berg, T.M., 2000, The National Geologic Map Database--A progress report, *in* Soller, D.R., editor, Digital Mapping Techniques '00—Workshop proceedings: U.S. Geological Survey Open-file Report 00-325, p. 27-30, <http://pubs.usgs.gov/openfile/of00-325/soller2.html>.

Soller, D.R., and Berg, T.M., 1999a, Building the National Geologic Map Database: Progress and challenges, *in* Derksen, C.R.M, and Manson, C.J., editors, Accreting the continent's collections: Geoscience Information Society Proceedings, v. 29, p. 47-55, <http://ncgmp.usgs.gov/ngmdbproject/reports/gisproc98.html>.

Soller, D.R., and Berg, T.M., 1999b, The National Geologic Map Database—A progress report, *in* Soller, D.R., editor, Digital Mapping Techniques '99—Workshop proceedings: U.S. Geological Survey Open-file Report 99-386, p. 31-34, <http://pubs.usgs.gov/openfile/of99-386/soller1.html>.

Soller, D.R., and Berg, T.M., 1998, Progress Toward Development of the National Geologic Map Database, *in* Soller, D.R., editor, Digital Mapping Techniques '98—Workshop proceedings: U.S. Geological Survey Open-file Report 98-487, p. 37-39, <http://pubs.usgs.gov/openfile/of98-487/soller2.html>.

Soller, D.R., and Berg. T.M., 1997, The National Geologic Map Database—A progress report: Geotimes, v. 42, no. 12, p. 29-31.

Soller, D.R., and Berg, T.M., 1995, Developing the National Geologic Map Database: Geotimes, v. 40, no. 6, p. 16-18.

APPENDIX C

List of publishers contained in the National Geologic Map Databases's Geoscience Map Catalog

Alabama Academy of Science
Alaska Division of Geological & Geophysical Surveys (1972-present)
Alaska Division of Geological Survey (1970-72)
Alaska Division of Mines and Geology (1966-70)
Alaska Division of Mines and Minerals (1959-66)
Alaska Territorial Department of Mines (1959)
American Association of Petroleum Geologists
American Geophysical Union
American Institute of Mining, Metallurgical, and Petroleum Engineers
Arizona Bureau of Geology and Mineral Technology
Arizona Bureau of Mines
Arizona Department of Water Resources
Arizona Geological Society
Arizona Geological Survey
Arizona Public Service
Arizona State University
Arkansas Geological Commission
Association of Engineering Geologists
Baylor University
Bowling Green State University
Brigham Young University Department of Geology
British Columbia Ministry of Energy and Mines
California Division of Mines and Geology
California Institute of Technology
California State University, Chico
California State University, Fresno
California State University, Humboldt
California State University, Long Beach
Canadian Hydrographic Service, Department of Fisheries and Oceans
Canyonlands Natural History Association
Colorado Geological Survey
Colorado School of Mines
Colorado State University
Columbia University Libraries
Columbia University School of Mines
Commonwealth of Virginia Department of Conservation and Economic Development
Confederated Tribes of the Colville Reservation
Connecticut Geological and Natural History Survey
Dallas Geological Society
Delaware Geological Survey
Desert Research Institute
Dibblee Geological Foundation
Eastern Washington University
Elsevier Science
Environment Canada
Field Conference of Pennsylvania Geologists, Inc.
Florida Geological Survey
Geodata International, Inc.
Geological Society of America
Geological Society of Nevada
Geological Society of Sacramento
Geological Survey Department, Jamaica
Geological Survey of Alabama
Geological Survey of Canada
Geological Survey of Michigan
Georgia Department of Natural Resources
Georgia Division of Mines, Mining, and Geology
Global Tectonics and Metallogeny
Grand Canyon Association
Great Plains Historical Association
GTR Mapping
Hawaii Commission on Water Resource Management
Hawaii Division of Water and Land Development
Hawaii Institute of Geophysics and Planetology
Hawaii Water Authority
Idaho Bureau of Mines and Geology
Idaho Geological Survey
Idaho State University
Illinois Basin Consortium
Illinois Oil and Gas Association
Illinois State Geological Survey
Indiana Department of Conservation
Indiana Department of Natural Resources
Indiana Geological Survey
Indiana University, Department of Geological Sciences
Institute of Food And Agricultural Sciences Service, University of Florida
Intergovernmental Resource Center, Clark County, Washington
Intermountain Association of Petroleum Geologists
IntraSearch, Inc
Iowa Geological Survey
John Wiley and Sons Publishers
Joint Transportation Research Program, Purdue University/Indiana Department of Transportation
Kansas Academy of Science
Kansas Geological Society
Kansas Geological Survey
Kentucky Department of Commerce
Kentucky Geological Survey
Lincoln-DeVore Engineers and Geologists
Loma Linda University
Los Alamos National Laboratory
Louisiana Geological Survey
Mackay School of Mines
Maine Geological Survey
Martel Laboratories, Inc.
Maryland Geological Survey

Massachusetts Institute of Technology
Medical Association of the State of Alabama
Memorial University of Newfoundland
Miami Geological Society
Miami University, Ohio
Michigan Department of Conservation
Michigan Department of Natural Resources
Mineral Resources Development, Inc.
Mines and Minerals (Scranton, PA)
Minnesota Department of Natural Resources, Division of Waters
Minnesota Geological Survey
Missouri Division of Geology and Land Survey
Missouri Geological Survey
Missouri Geological Survey and Resource Assessment Division
Montana Bureau of Mines and Geology
Montclaire State College, NJ
Mountain Press Publishing Company
Museum of Northern Arizona
National Academy of Sciences - National Research Council
National Well Water Association
Nevada Bureau of Mines and Geology
Nevada Department of Conservation and Natural Resources
Nevada Division of Water Resources
Nevada Petroleum Society
New Hampshire Academy of Science
New Hampshire Department of Environmental Services
New Hampshire Department of Resources & Economic Development
New Hampshire State Planning and Development Commission
New Jersey Geological Survey
New Mexico Bureau of Geology and Mineral Resources
New Mexico Bureau of Mines and Mineral Resources
New Mexico Geological Society
New York Academy of Sciences
New York State Department of Environmental Conservation
New York State Geological Survey
New York State Museum
New York, Oswego County Planning Board
North Carolina Department of Natural Resources and Community Development
North Carolina Department of Transportation, Geotechnical Engineering Unit
North Carolina Division of Mineral Resources
North Carolina Geological Survey
North Dakota Geological Survey
Northern Arizona University
Northwest Scientific Association
Northwestern University
Ohio Division of Geological Survey
Ohio Division of Shore Erosion (Ohio Division of Geological Survey)
Ohio Geological Society
Ohio State Univesity
Ohio University
Oklahoma Geological Survey
Oklahoma State University
Oregon Department of Geology and Mineral Industries
Oregon State University
Oxford University Press
Paleontological Research Institution
Pennsylvania First Geological Survey (1836-1842)
Pennsylvania Geological Survey
Pennsylvania Second Geological Survey (1874-1889)
Pennsylvania State University
Pennsylvania Third Geological Survey (1899-1914)
Petroleum Publishing Company
Portland State University Department of Geology
Primedia Business Magazines & Media
Princeton University
Puerto Rico Department of Public Works
Puerto Rico Division of Mineralogy and Geology
Puget Sound Power and Light Company
Purdue University
Purdue University Office of Agricultural Research Programs
Rhode Island Geological Survey
Rice University
Rockwell International, Rockwell Hanford Operations, Energy System Group
Royal Bank of Canada, Oil and Gas Department
San Diego State University
San Jose State University
Shannon & Wilson, Inc
Sigma Gamma Epsilon
Society of Economic Geologists
Society of Economic Paleontologists and Mineralogists
South Carolina Geological Survey
South Coast Geological Society, Inc.
South Dakota Academy of Science
South Dakota Geological Survey
Southern California Academy of Sciences
Southern Pacific Railroad
Springer-Verlag New York
Stanford University
State Geological Survey of Kansas
State of New Jersey Department of Conservation and Economic Development
Tacoma-Pierce County Health Department
Tennessee Division of Geology
Terrascan Group Ltd., Lakewood, CO
Texas A&M University
Texas Christian University
Texas Tech University
TRW, Inc
Tulane University
U.S. Army Corps of Engineers
U.S. Atomic Energy Commission

U.S. Bureau of Mines
U.S. Bureau of Reclamation
U.S. Department of Agriculture, Forest Service
U.S. Department of Agriculture, Natural Resources Conservation Service
U.S. Department of Energy
U.S. Department of Energy, Grand Junction Office
U.S. Department of Energy, Morgantown Energy Technology Center
U.S. Department of Transportation, Federal Highway Administration, Indiana Division
U.S. Geological Survey
U.S. National Oceanic and Atmospheric Administration
University of Alabama
University of Alaska, Fairbanks
University of Arizona
University of Arizona, Department of Geosciences
University of Arkansas
University of California
University of California, Davis
University of California, Los Angeles
University of California, Riverside
University of California, Santa Barbara
University of Chicago Press
University of Colorado, Boulder
University of Hawaii, Water Resources Research Center
University of Idaho
University of Illinois, Urbana-Champaign
University of Iowa
University of London
University of Missouri, Columbia
University of Missouri, Rolla
University of Nebraska Conservation and Survey Division
University of Nebraska, Lincoln
University of Nevada Las Vegas
University of Nevada, Reno
University of New Mexico
University of New Orleans
University of North Carolina at Chapel Hill
University of Oklahoma
University of Oregon
University of Puerto Rico
University of South Carolina
University of Texas at Austin, Bureau of Economic Geology
University of Texas, Austin
University of Texas, El Paso
University of Toledo
University of Tulsa
University of Utah
University of Utah Research Institute, Earth Science Laboratory Research Institute
University of Washington
University of Wisconsin
University of Wisconsin, Madison
University of Wisconsin, Milwaukee
University of Wyoming
Utah Department of Natural Resources and Energy
Utah Geological and Mineral Survey
Utah Geological and Mineralogical Survey
Utah Geological Association
Utah Geological Survey
Vermont Department of Water Resource
Vermont Geological Survey
Virginia Division of Mineral Resources
Washington Department of Conservation and Development
Washington Department of Ecology
Washington Division of Geology and Earth Resources
Washington Division of Mines and Geology
Washington Division of Water Resources
Washington Geological Survey
Washington State University
West Virginia Geological and Economic Survey
West Virginia Geological Survey
Western Michigan University Department of Geology
Willard Owens Associates, Inc.
Wisconsin Geological and Natural History Survey
Wright State University
Wyoming Geological Association
Wyoming State Geological Survey
Yale University

CHRONOS—Integrated Stratigraphic Databases, the Development of an International Standard Time Scale and the Interoperability with Time Scales of U.S. State Surveys

By Bruce R. Wardlaw

U.S. Geological Survey
926A National Center
Reston, VA 20192
Telephone: 703-648-5288
Email: bwardlaw@usgs.gov

INTRODUCTION

Modern Earth system history research increasingly depends upon the analysis of voluminous multidisciplinary, time-calibrated data. The process of determining the availability or even the existence of Earth history data remains a time-consuming and error-prone enterprise because there are no centralized depositories or Web-enabled means for locating and retrieving data. The goal of CHRONOS is to unify current and future stratigraphic databases into a powerful system for producing a dynamic global time scale for Earth history and for understanding the complex relationships of past geologic, climatic and evolutionary trends. Web-services and an extensive on-line suite of toolboxes will allow global researchers and the general public to access, analyze and visualize CHRONOS stratigraphic information. Another goal, and the centerpiece of the CHRONOS system, is a vastly improved high-resolution geological time scale.

A logical series of tasks is planned to accomplish these goals, with the ultimate goal being for all geoscientists to be able to apply the CHRONOS system of integrated databases for deciphering the complex interactions of the Earth system through all of geologic time. In addition to its primary goals of enabling networking of international databases and facilitating creative research, the CHRONOS outreach programs will promote education of Earth's fascinating history.

CHRONOS is a system within the larger effort of the National Science Foundation for cyberinfrastructure in the earth sciences (called Geoinformatics). It will be a distributed system with a single portal at a central hub that will provide information technology expertise and tools and link the various thematic nodes or networks, projects, and databases constituting CHRONOS (fig. 1).

The implementation of CHRONOS will involve the following tasks:

- Progressively establish and interlink critical thematic database networks for life through time, radiometric age, climate through time, chemical and sedimentary cycles through time, core-to-crust dynamics through time (such as magnetostratigraphy), and other components of Earth system history;
- Establish central CHRONOS portal to access and analyze major component data types for researchers and the general public;
- Develop advanced tools and visualization capabilities that investigators can apply to uploaded data in their on space on a CHRONOS server or download to their own system;
- Assemble a high-precision 'Standard Geological Time Scale' under the aegis of the International Commission on Stratigraphy;
- Coordinate an outreach program with educational modules and informative demonstrations of the CHRONOS system; and
- Study four critical time-slices of Earth history as 'test-bed' investigations, using the expanding capabilities of the information-technology infrastructure and toolkits of the assembled CHRONOS system.

The four time-slice studies address longstanding scientific questions of societal relevance and interest (Cambrian life explosion, Permian-Triassic catastrophic extinctions, middle Cretaceous super-greenhouse world, and middle Miocene climate transitions; see fig. 2). Each involves different types and qualities of data and will improve and refine marine and continental data cor-

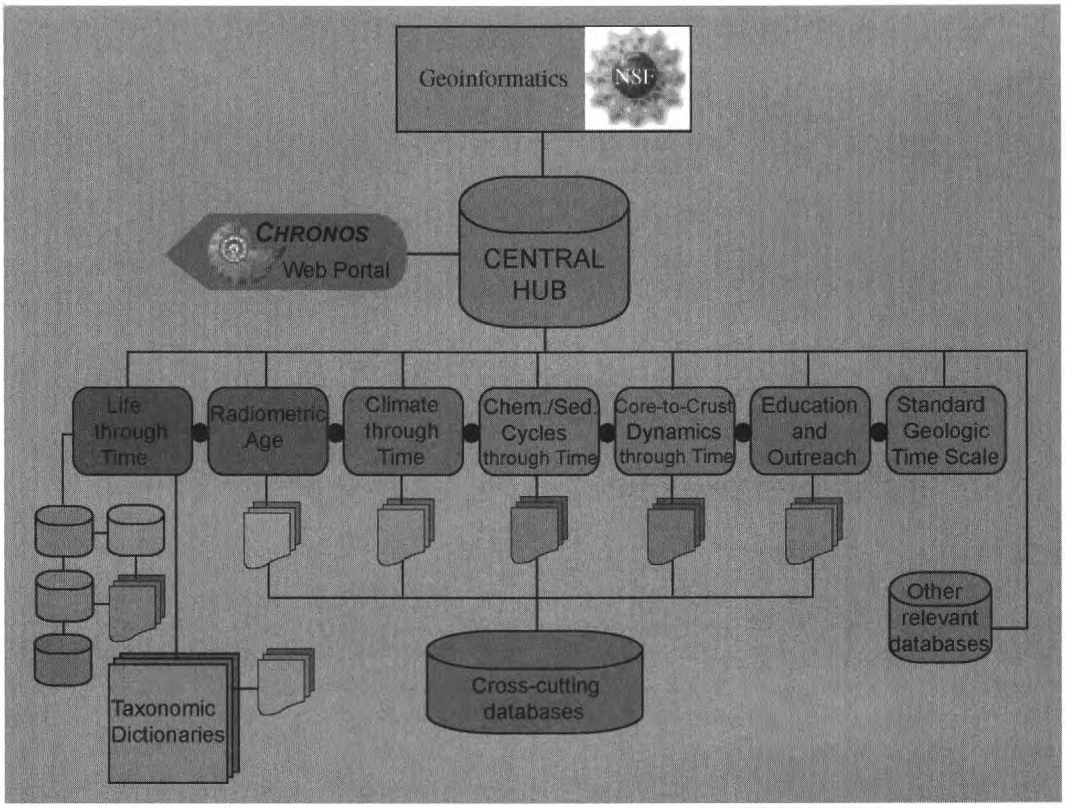
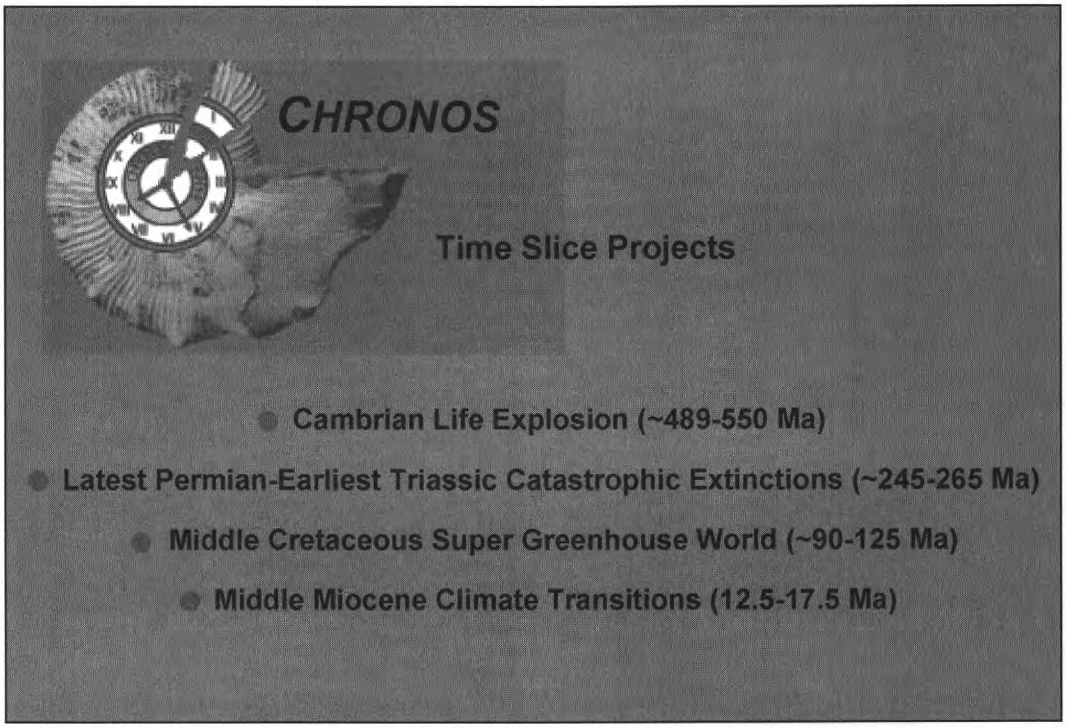


Figure 1. Schematic of the distributed CHRONOS system.



CHRONOS

Time Slice Projects

- Cambrian Life Explosion (~489-550 Ma)
- Latest Permian-Earliest Triassic Catastrophic Extinctions (~245-265 Ma)
- Middle Cretaceous Super Greenhouse World (~90-125 Ma)
- Middle Miocene Climate Transitions (12.5-17.5 Ma)

Figure 2. Time Slice Projects of CHRONOS.

relations at the global scale. The ultimate goal is for all geoscientists to be able to apply the CHRONOS system of integrated databases for deciphering the complex interactions of the Earth system through all of geologic time.

HOW THE CHRONOS INTERACTIVE NETWORK WILL WORK WITH DMT AND THE U.S STATE SURVEYS

An example of how CHRONOS will support the goals of the Digital Mapping Techniques workshops is the seamless way it will relate local state geological survey time scales to the International Time Scale. Initial information and links for both local and international scales will be available from the National Geologic Map Database's Geological Names Lexicon ("GEOLEX", http://ngmsvr.wr.usgs.gov/Geolex/). The current international scale is available from the International Commission on Stratigraphy (ICS, http://www.micropress.org/stratigraphy/index.htm) or a slightly modified version (retaining the Tertiary) from the U. S. Geological Survey (USGS), soon to be available on GEOLEX (fig. 3).

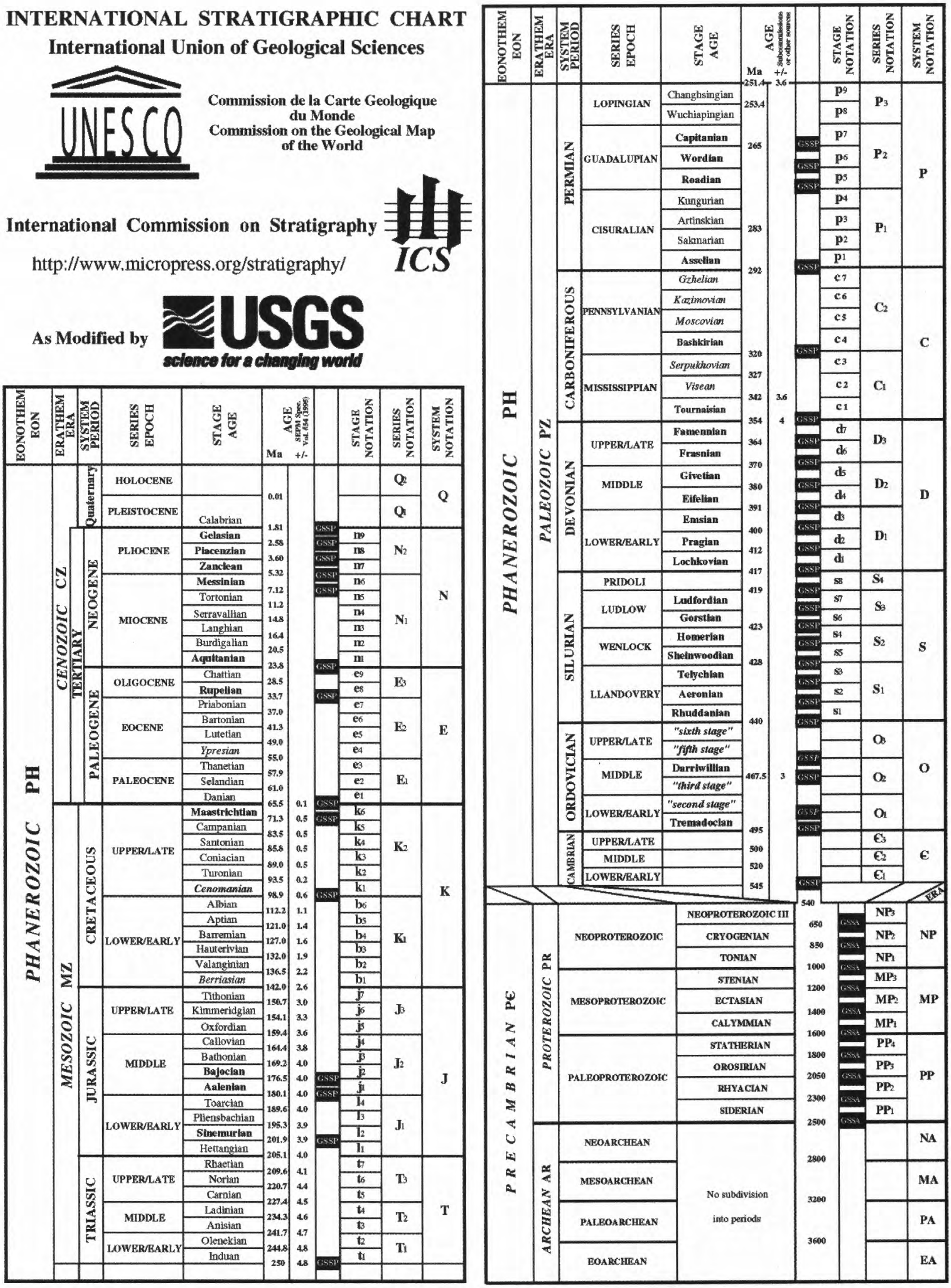
As an example of how CHRONOS will function, let us imagine that we want to know where the currently internationally-defined Carboniferous-Permian boundary falls in the sections in Kansas and West Texas and how it relates to the boundary recognized by those states. From the CHRONOS portal, you would be linked to the ICS site containing the Global Stratotype Section and Points (GSSP's) for all formally accepted boundaries of the ICS (currently, go to http://www.micropress.org/stratigraphy/index.htm, and select "GSSP's"). The Carboniferous-Permian boundary information displayed is as follows:

base Asselian Stage, base Cisuralian Series, base Permian System	Conodont, lowest occurrence of *Streptognathodus isolatus* within the *S. "wabaunsensis"* conodont chronocline. 6 m higher is lowest fusilinid foraminifer *Sphaeroschwagerina*	27 m above base of Bed 19, Aidaralash Creek, Aktöbe, southern Ural Mountains, northern Kazakhstan	Ratified, 1996	Episodes 21 (1), p.11

The first box leads to: http://www.micropress.org/stratigraphy/carper.htm where we find full references and section and map; the *Episodes* article also is available digitally.

From a completely different source identified by CHRONOS, we would get a current reinterpretation of the conodont lineages from the GSSP (http://www.kgs.ukans.edu/, fig. 4), as found in Wardlaw and others (2003). From the same publication, the conodont ranges from Kansas are shown (fig. 5). A brief history of the placement of the Permian boundary in Kansas would be available from Paleodata (the National Geologic Map Database's National Paleontologic Database, fig. 6) which will show where R. C. Moore placed the boundary at the top of the Brownville Limestone, which became the "traditional" boundary; and where the boundary was modified using the first occurrence of the *Streptognathodus constrictus* conodont zone (an early contender for the boundary definition) placed by Baars and others (1994) at the base of the Neva Limestone. However, Boardman and others (1998) clearly show the first occurrence of *S. constrictus* and other species that make up the zone in the underlying Burr Limestone. The first occurrence of *S. isolatus* (the international boundary indicator) is in the Bennett Shale of the Red Eagle Limestone.

Further, in the reference section for the Carboniferous-Permian boundary at Usolka in the Urals, Russia, there are found volcanic tuffs centimeters below and above the first occurrence of *Streptognathodus isolatus* that yield abundant zircons (*Permophiles* no. 39, http://pri.boisestate.edu/permophiles/issue39.pdf). Both horizons indicate virtually the same date, and though preliminary, are dated at 299 million years before present(Ma). This provides a solid age that can be applied to both Kansas and West Texas sections. Again, from the ICS



INTERNATIONAL STRATIGRAPHIC CHART
International Union of Geological Sciences

Commission de la Carte Geologique du Monde
Commission on the Geological Map of the World

International Commission on Stratigraphy

http://www.micropress.org/stratigraphy/

As Modified by USGS

EONOTHEM EON	ERATHEM ERA	SYSTEM PERIOD	SERIES EPOCH	STAGE AGE	AGE Ma	SEPM Spec. Vol. #54 (1995) +/-		STAGE NOTATION	SERIES NOTATION	SYSTEM NOTATION
PHANEROZOIC PH	CENOZOIC CZ	Quaternary	HOLOCENE		0.01				Q_2	Q
			PLEISTOCENE	Calabrian	1.81		GSSP		Q_1	
		TERTIARY NEOGENE	PLIOCENE	**Gelasian**	2.58		GSSP	n_9	N_2	N
				Piacenzian	3.60		GSSP	n_8		
				Zanclean	5.32		GSSP	n_7		
			MIOCENE	**Messinian**	7.12		GSSP	n_6	N_1	
				Tortonian	11.2			n_5		
				Serravallian	14.8			n_4		
				Langhian	16.4			n_3		
				Burdigalian	20.5			n_2		
				Aquitanian	23.8		GSSP	n_1		
		PALEOGENE	OLIGOCENE	Chattian	28.5			e_9	E_3	E
				Rupelian	33.7		GSSP	e_8		
			EOCENE	Priabonian	37.0			e_7	E_2	
				Bartonian	41.3			e_6		
				Lutetian	49.0			e_5		
				Ypresian	55.0			e_4		
			PALEOCENE	Thanetian	57.9			e_3	E_1	
				Selandian	61.0			e_2		
				Danian	65.5	0.1	GSSP	e_1		
	MESOZOIC MZ	CRETACEOUS	UPPER/LATE	**Maastrichtian**	71.3	0.5	GSSP	k_6	K_2	K
				Campanian	83.5	0.5		k_5		
				Santonian	85.8	0.5		k_4		
				Coniacian	89.0	0.5		k_3		
				Turonian	93.5	0.2		k_2		
				Cenomanian	98.9	0.6	GSSP	k_1		
			LOWER/EARLY	Albian	112.2	1.1		b_6	K_1	
				Aptian	121.0	1.4		b_5		
				Barremian	127.0	1.6		b_4		
				Hauterivian	132.0	1.9		b_3		
				Valanginian	136.5	2.2		b_2		
				Berriasian	142.0	2.6		b_1		
		JURASSIC	UPPER/LATE	Tithonian	150.7	3.0		j_7	J_3	J
				Kimmeridgian	154.1	3.3		j_6		
				Oxfordian	159.4	3.6		j_5		
			MIDDLE	Callovian	164.4	3.8		j_4	J_2	
				Bathonian	169.2	4.0		j_3		
				Bajocian	176.5	4.0	GSSP	j_2		
				Aalenian	180.1	4.0	GSSP	j_1		
			LOWER/EARLY	Toarcian	189.6	4.0		l_4	J_1	
				Pliensbachian	195.3	3.9		l_3		
				Sinemurian	201.9	3.9	GSSP	l_2		
				Hettangian	205.1	4.0		l_1		
		TRIASSIC	UPPER/LATE	Rhaetian	209.6	4.1		t_7	T_3	T
				Norian	220.7	4.4		t_6		
				Carnian	227.4	4.5		t_5		
			MIDDLE	Ladinian	234.3	4.6		t_4	T_2	
				Anisian	241.7	4.7		t_3		
			LOWER/EARLY	Olenekian	244.8	4.8		t_2	T_1	
				Induan	250	4.8	GSSP	t_1		

EONOTHEM EON	ERATHEM ERA	SYSTEM PERIOD	SERIES EPOCH	STAGE AGE	AGE Ma	Subcommissions or other sources +/-		STAGE NOTATION	SERIES NOTATION	SYSTEM NOTATION
					251.4	3.6				
PHANEROZOIC PH	PALEOZOIC PZ	PERMIAN	LOPINGIAN	Changhsingian	253.4			p_9	P_3	P
				Wuchiapingian				p_8		
			GUADALUPIAN	**Capitanian**	265		GSSP	p_7	P_2	
				Wordian			GSSP	p_6		
				Roadian			GSSP	p_5		
			CISURALIAN	Kungurian				p_4	P_1	
				Artinskian	283			p_3		
				Sakmarian				p_2		
				Asselian	292		GSSP	p_1		
		CARBONIFEROUS	PENNSYLVANIAN	*Gzhelian*				c_7	C_2	C
				Kazimovian				c_6		
				Moscovian				c_5		
				Bashkirian	320		GSSP	c_4		
			MISSISSIPPIAN	*Serpukhovian*	327			c_3	C_1	
				Visean	342	3.6		c_2		
				Tournaisian	354	4	GSSP	c_1		
		DEVONIAN	UPPER/LATE	**Famennian**	364		GSSP	d_7	D_3	D
				Frasnian	370		GSSP	d_6		
			MIDDLE	**Givetian**	380		GSSP	d_5	D_2	
				Eifelian	391		GSSP	d_4		
			LOWER/EARLY	**Emsian**	400		GSSP	d_3	D_1	
				Pragian	412		GSSP	d_2		
				Lochkovian	417		GSSP	d_1		
		SILURIAN	PRIDOLI		419		GSSP	s_8	S_4	S
			LUDLOW	**Ludfordian**			GSSP	s_7	S_3	
				Gorstian	423		GSSP	s_6		
			WENLOCK	**Homerian**			GSSP	s_4	S_2	
				Sheinwoodian	428		GSSP	s_5		
			LLANDOVERY	**Telychian**			GSSP	s_3	S_1	
				Aeronian			GSSP	s_2		
				Rhuddanian	440		GSSP	s_1		
		ORDOVICIAN	UPPER/LATE	***"sixth stage"***					O_3	O
				"fifth stage"			GSSP			
			MIDDLE	**Darriwillian**	467.5	3	GSSP		O_2	
				"third stage"						
			LOWER/EARLY	***"second stage"***			GSSP		O_1	
				Tremadocian	495		GSSP			
		CAMBRIAN	UPPER/LATE		500				$Є_3$	Є
			MIDDLE		520				$Є_2$	
			LOWER/EARLY		545		GSSP		$Є_1$	

							ERA
				540			
PRECAMBRIAN PЄ	PROTEROZOIC PR	NEOPROTEROZOIC	NEOPROTEROZOIC III	650	GSSA	NP_3	NP
			CRYOGENIAN	850	GSSA	NP_2	
			TONIAN	1000	GSSA	NP_1	
		MESOPROTEROZOIC	STENIAN	1200	GSSA	MP_3	MP
			ECTASIAN	1400	GSSA	MP_2	
			CALYMMIAN	1600	GSSA	MP_1	
		PALEOPROTEROZOIC	STATHERIAN	1800	GSSA	PP_4	PP
			OROSIRIAN	2050	GSSA	PP_3	
			RHYACIAN	2300	GSSA	PP_2	
			SIDERIAN	2500	GSSA	PP_1	
	ARCHEAN AR	NEOARCHEAN	No subdivision into periods	2800			NA
		MESOARCHEAN		3200			MA
		PALEOARCHEAN		3600			PA
		EOARCHEAN					EA

Figure 3. International Stratigraphic Chart as modified by USGS.

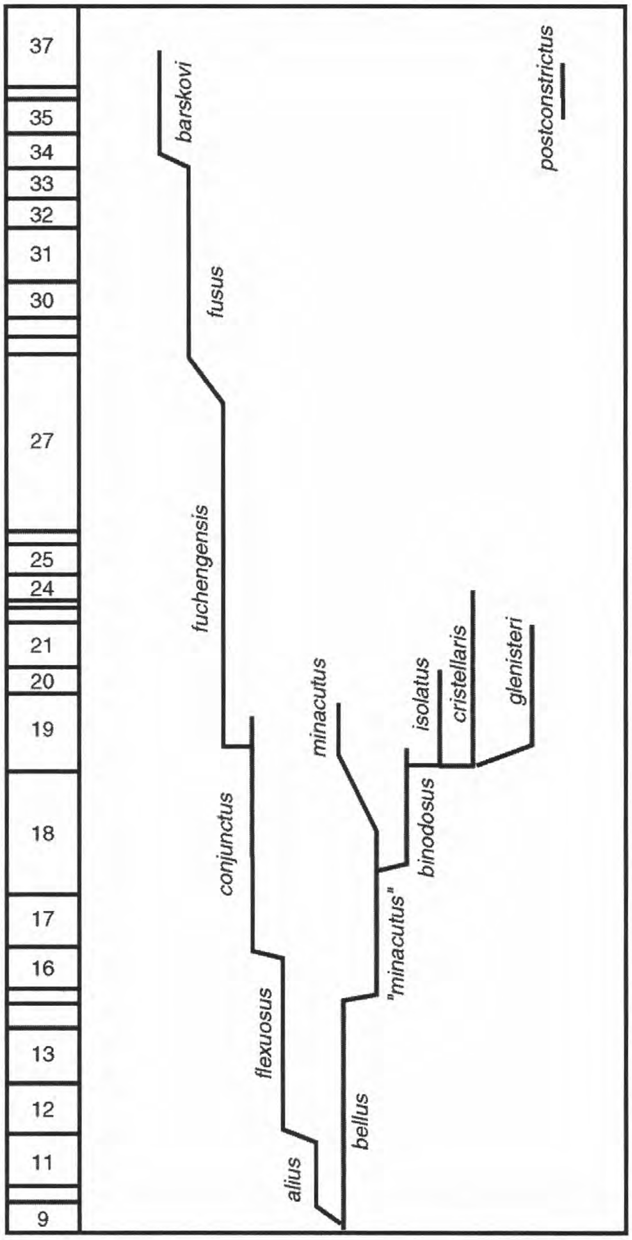


Figure 4. Interpretation of *Streptognathodus* conodont lineages from the GSSP of the Permian at Aidaralash Creek, Kazakhstan; section shown as numbered beds (from Wardlaw and others, 2003).

website, information on relevant radiometric dates utilized in constructing the Phanerozoic time scale is available (fig. 7, from Gradstein and Ogg, 2004, to be made available through ICS).

For the necessary stratigraphic information for West Texas, we would be linked to a digital copy of an article from *Permophiles* issue number 36 (June, 2000, http://pri.boisestate.edu/permophiles/issue36.pdf) which follows:

"Preliminary Placement of the International Permian Working Standard to the Glass Mountains, Texas"

Bruce R. Wardlaw, U.S. Geological Survey, Reston, VA
Vladimir I. Davydov, Permian Research Institute, Boise State Univ., Boise, ID

The International Lower Permian Working Standard is based on the conodont distributions in the southern Urals, Kansas, and West Texas. Asselian is defined on the first appearance of *Streptognathodus isolatus*. This occurs at the base of the Bennett Shale of the Red Eagle Limestone in Kansas. The current working definition for the base of the Sakmarian is the First Appearance (FA) of *Streptognathodus barskovi (sensu strictu)*. This occurs in the Eiss Limestone of the Bader Limestone in Kansas and is very close to the FA of *Sweetognathus merrilli* in the upper part of the Eiss. The working definition of the base of the Artinskian is the FA of *S. florensis* or *Sweetognthus whitei*, which both first appear in the base of the Florence Limestone of the Barnston Formation in Kansas. After Barneston deposition the Kansas section remains very shallow during marine incursions and only sparse to common *Sweetognathus* or *Rabegnathus* faunas are recovered. The working standard for Kungurian conodont zonation is based on the distribution of conodonts from the Glass Mountains, Texas. The working definition of the base of the Kungurian is the FA of *Neostreptognathodus "exsculptus"*.

The Grey Limestone Member of the Gaptank Formation is a shallow-water carbonate that forms the top of the formation. Conodont faunas are sparse, but the base of the Grey Limestone Member contains a conodont fauna that correlates to the Foraker Limestone with the overlap of the ending range of *Streptognathodus brownvillensis* and *S. elongatus*. The top of the Grey Limestone contains a conodont fauna that correlates to the Grenola Limestone with *Streptognathodus nevaensis* and *S. elongatus*. Therefore, the Carboniferous-Permian boundary is located within the Grey Limestone.

First Appearance of	Stratigraphic Position
N. "exsculptus"	17 m above base, Skinner Ranch Formation
S. barskovi	52 m above base, Neal Ranch Formation
S. nevaensis	Top of Grey Limestone Member
S. elongatus	Base of Grey Limestone Member

The Neal Ranch Formation is dominated by prodelta siltstones with common plant debris and limestone and limestone conglomerate interbeds. The limestones yield a fair conodont fauna. At 52 m above the base *Streptognathodus barskovi, S. isolatus* and *Sweetongnathus merrilli* occur (equivalent to bed 12 of Ross, 1963). The overlap of *barskovi* and *isolatus* indicates the base of the *S. barskovi* zone. *Streptognathodus barskovi* occurs higher at 71 m where it occurs with *S. postconstrictus* which indicates the upper part of its zone.

The Lenox Hills Formation is largely delta conglomerates and does not yield conodonts. However, in the Dugout Mountain area, it contains common limestone which yields a sparse fauna of *Sweetognathus whitei* and *Neostreptognathodus transitus*. This fauna indicates the upper Artinskian. Conodonts do not well constrain the Sakmarian-Artinskian boundary.

The Skinner Ranch formation is largely limestone and limestone conglomerate with abundant conodont fuanas. In its type section it contains *Neostreptognathodus pequopensis* at its base. At 17 m, a plethora of species of *Mesogondolella* and *Streptognathodus* occur, including *Neostreptognathodus "exsculptus"* which marks the base of the Kungurian."

Since publication of that article, the working definitions have been refined and changed to formal proposals, and the definition of the Sakmarian has changed from the First Appearance Datum (FAD) of *Streptognathodus barskovi* to the FAD of *Sweetognathus merrilli* (*Permophiles* no. 41, http://pri.boisestate.edu/permophiles/issue41.pdf)

We want more substantial information, so CHRONOS would conduct an author search for Wardlaw and for Davydov, and find that there is no further information published by these authors on the placement of conodonts in the lower part of the West Texas regional stratotype. However, in our author search, we find all of Wardlaw's field information is available via PaleoStrat (https://www.paleostrat.com), a paleontologic and stratigraphic information system that allows us to see the measured section (fig. 8), information on each sample, and the conodont faunas recovered and digital images of the representative specimens. Thus, we are assured of the placement of the Carboniferous-Permian boundary in the Wolfcamp Hills/Geologist Canyon section, the regional stratotype of the Wolfcampian.

CHRONOS will bring all this disparate information together for the user. The sources of information also will be accessible, so that additional ideas and questions can be posed by the user.

ACKNOWLEDGMENTS

Much of the description of CHRONOS comes from the interaction of the Steering Committee and their work on writing the NSF proposal (see http://www.CHRONOS.org). The author was only one of many participants and is indebted to all of CHRONOS' creators. The author is solely responsible for the time scale interpretations. I thank David Soller for a rigorous review of the article.

REFERENCES

Baars, D. L., Ross, C. A, Ritter, S. M., and Maples, C. G., 1994, Proposed repositioning of the Pennsylvanian-Permian boundary in Kansas, *in* Baars, D. L., compiler, Revision of stratigraphic nomenclature in Kansas: Kansas Geological Survey Bulletin 230, p. 5-10.

Boardman, D. R., II, Nestell, M. K., and Wardlaw, B. R., 1998, Uppermost Carboniferous and lowermost Permian deposition and conodont biostratigraphy of Kansas, USA, *in* Jin, Y. G., Wardlaw, B. R., and Wang, Yue, eds., Permian stratigraphy, environments and resources, vol. 2, Stratigraphy and environments: Palaeoworld, no. 9, p. 19-32.

Gradstein, F. M., and Ogg, J. G., editors, 2004, A geologic time scale 2004: Cambridge University Press (in press). .

Ross, C.A., 1963, Standard Wolfcampian Series (Permian), Glass Mountains, Texas: Geological Society of America Memoir 88, 205p.

Wardlaw, B. R., Boardman, D. R., II, and Nestell, M. K., 2003, Conodont biostratigraphy of the uppermost Carboniferous and Lower Permian of Kansas: Kansas Geological Survey Bulletin (in press).

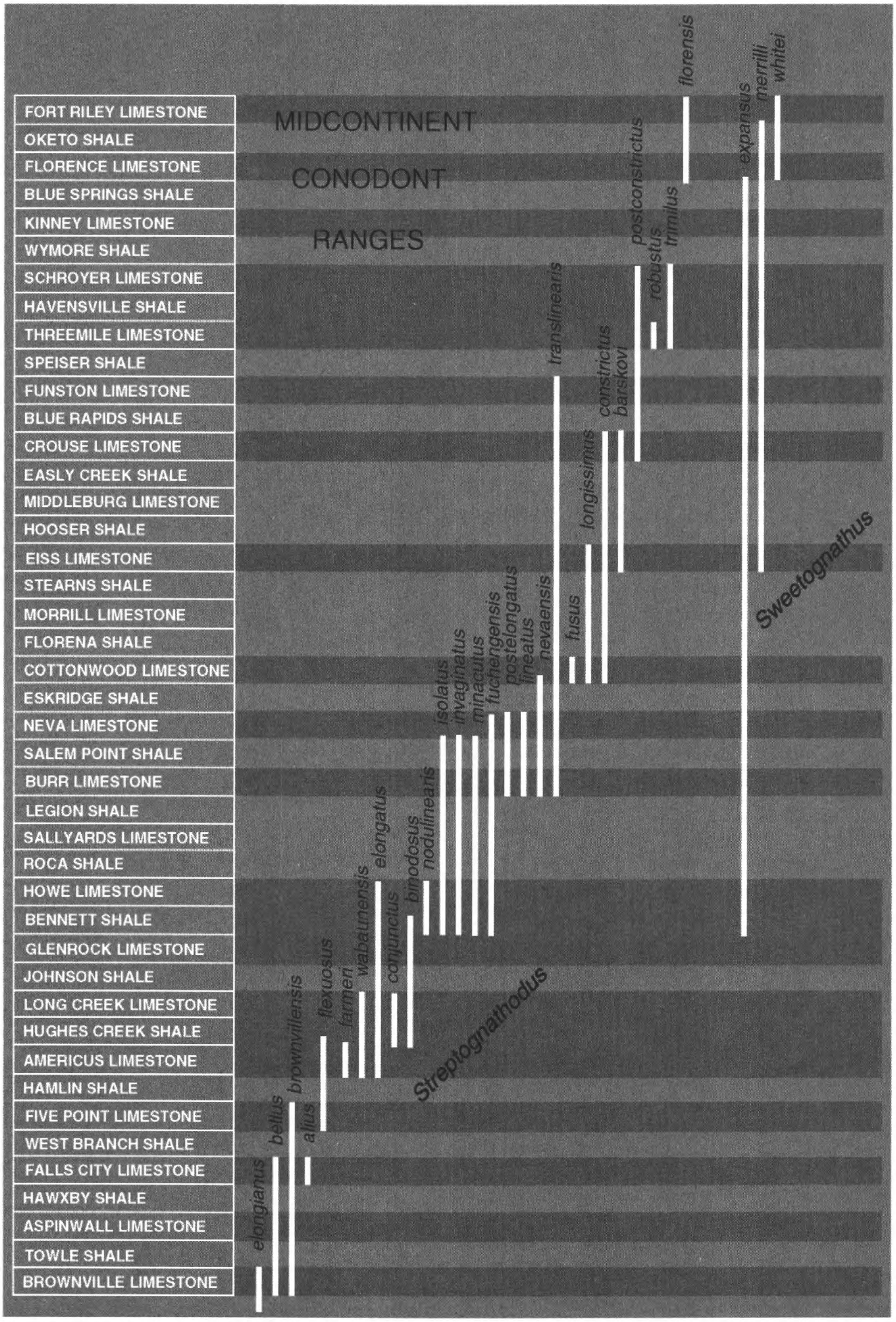
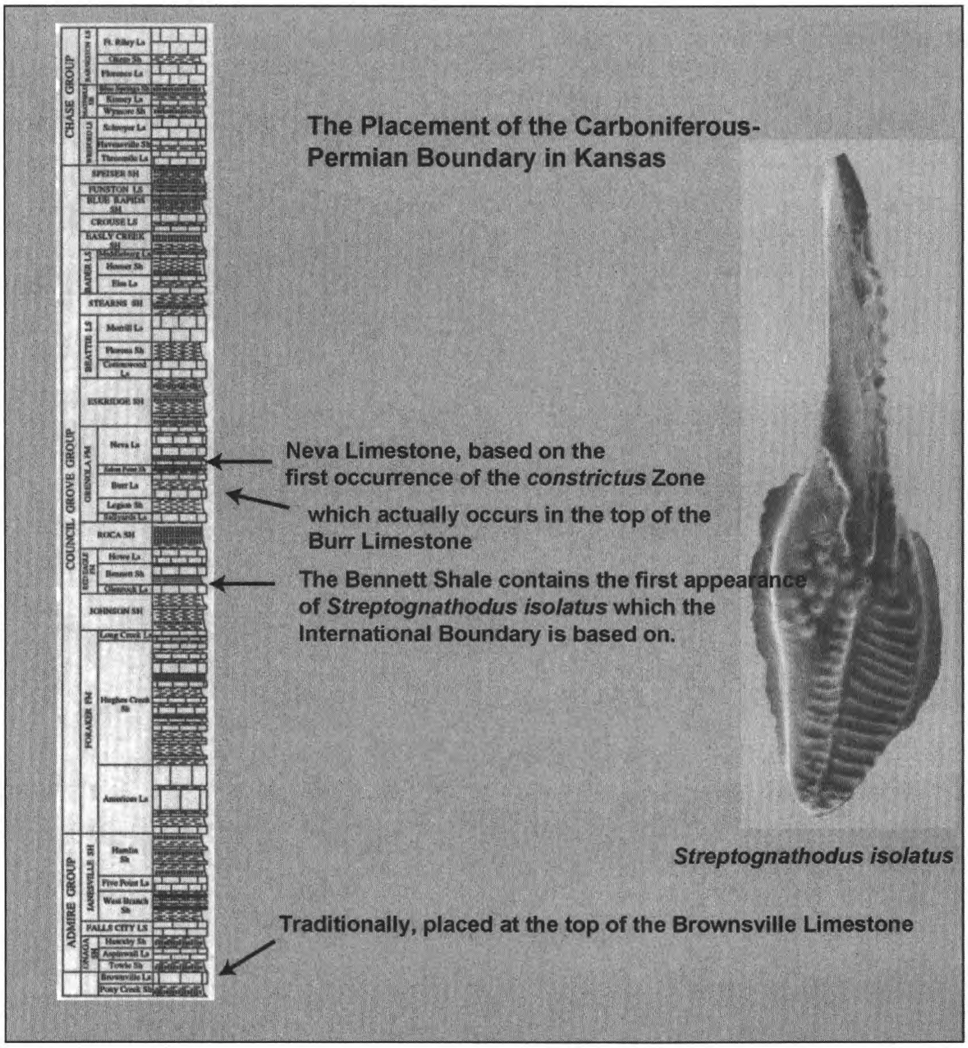


Figure 5. Ranges of conodont species in the Kansas section; darker intervals reflect largely marine deposition, lighter intervals reflect largely non- or marginal marine deposition (from Wardlaw and others, 2003).



The Placement of the Carboniferous-Permian Boundary in Kansas

Neva Limestone, based on the first occurrence of the *constrictus* Zone

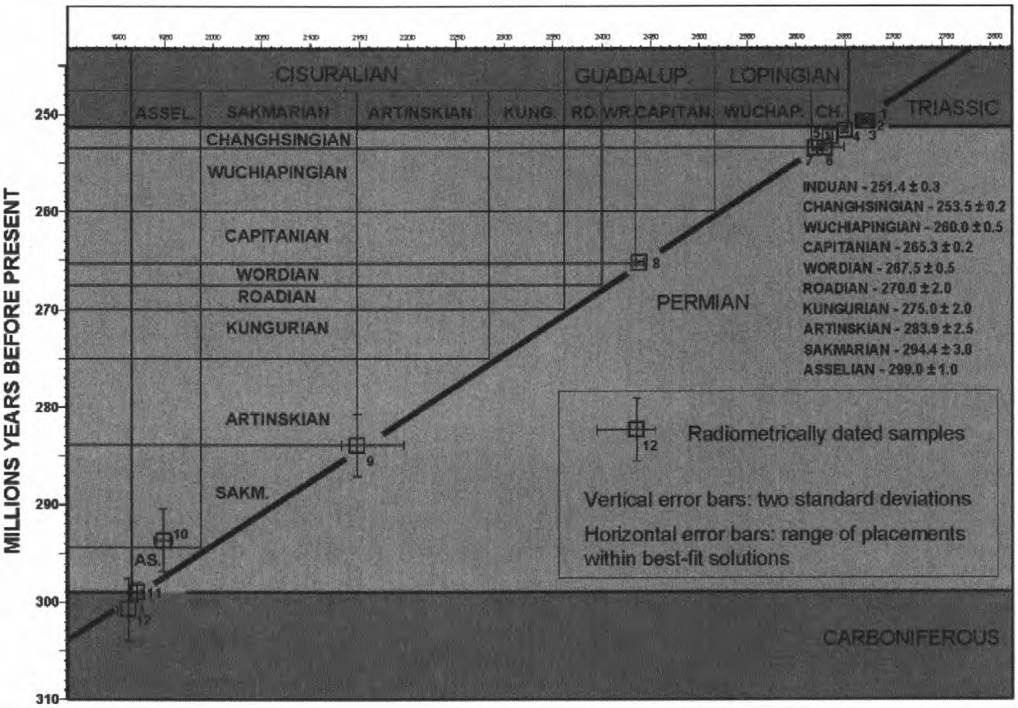
which actually occurs in the top of the Burr Limestone

The Bennett Shale contains the first appearance of *Streptognathodus isolatus* which the International Boundary is based on.

Streptognathodus isolatus

Traditionally, placed at the top of the Brownsville Limestone

Figure 6. Historical Permian boundary placement in the Kansas section.



INDUAN - 251.4 ± 0.3
CHANGHSINGIAN - 253.5 ± 0.2
WUCHIAPINGIAN - 260.0 ± 0.5
CAPITANIAN - 265.3 ± 0.2
WORDIAN - 267.5 ± 0.5
ROADIAN - 270.0 ± 2.0
KUNGURIAN - 275.0 ± 2.0
ARTINSKIAN - 283.9 ± 2.5
SAKMARIAN - 294.4 ± 3.0
ASSELIAN - 299.0 ± 1.0

Radiometrically dated samples

Vertical error bars: two standard deviations

Horizontal error bars: range of placements within best-fit solutions

Figure 7. Graphical correlation of radiometric ages and stage boundaries of the Permian.

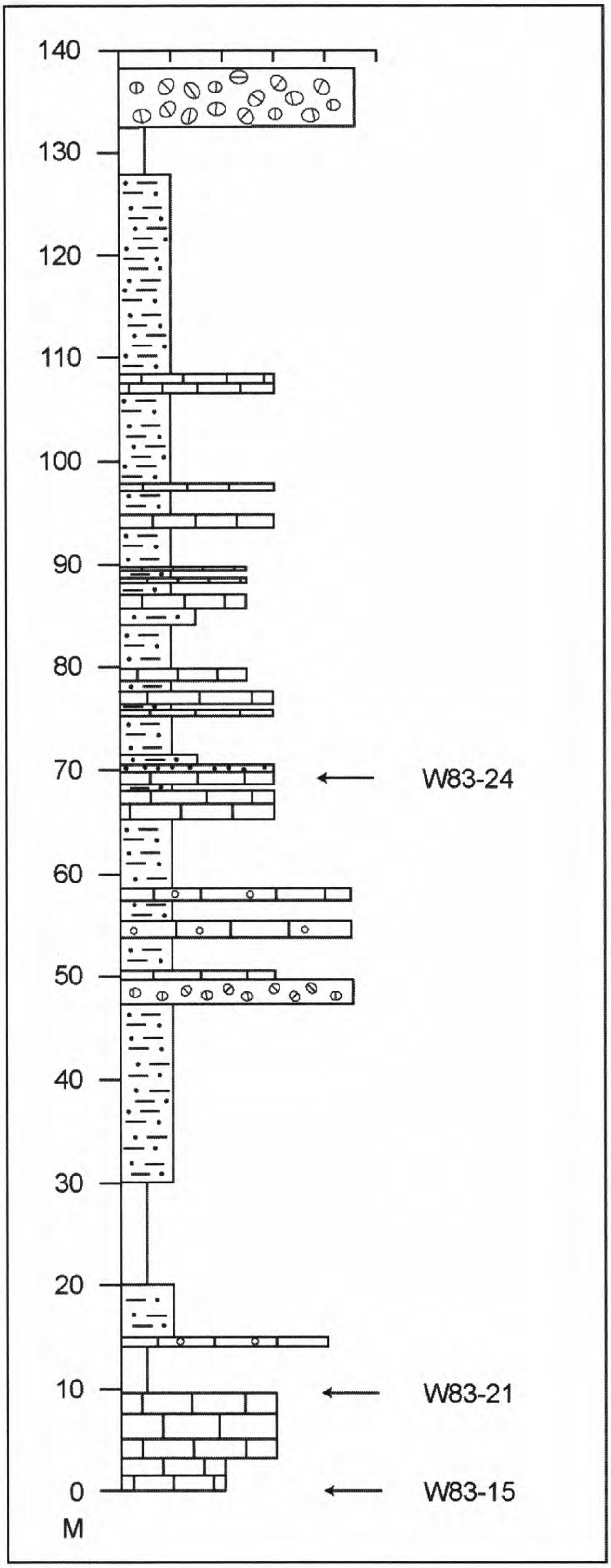


Figure 8. Columnar section of Wolfcamp Hills/Geologist Canyon section of Wardlaw and others (2003), and position of significant conodont samples. Measured in 1983 and recollected in 1984 (from PaleoStrat, definition of horizontal scale based on grain size also in PaleoStrat).

The National Geologic Map Database Image Library—General Concepts

By David R. Soller[1] and Thomas M. Berg[2]

[1]U.S. Geological Survey
926-A National Center
Reston, VA 20192
Telephone: (703) 648-6907
Fax: (703) 648-6977
e-mail: drsoller@usgs.gov

[2]Ohio Geological Survey
Telephone: (614) 265-6988
Fax: (614) 268-3669
e-mail: thomas.berg@dnr.state.oh.us

INTRODUCTION

The Geologic Mapping Act of 1992 and its reauthorizations mandate creation of a National Geologic Map Database (NGMDB), to serve as a "national archive" of geoscience maps. The U.S. Geological Survey (USGS) and the state geological surveys (through the Association of American State Geologists, or AASG) are collaborating on the design and the many components of the NGMDB project.

Consistent with the Act's intention, the project's overall objective is to provide the public with information and access to products that may be needed for purposes ranging from decisionmaking to education and personal interest. The mandate is at least partially fulfilled through the NGMDB's Map Catalog (Soller and Berg, this volume; Soller and Berg, 1999). The Map Catalog (<http://ngmsvr.wr.usgs.gov/ngmdb/ngm_catalog.ora.html>) lists bibliographic information about each geologic map and other geoscience products published for the United States and its territories and possessions. The bibliographic information and the contact information provided by the Map Catalog facilitate the borrowing, purchasing, and, in some cases, the downloading of map information.

We also are designing a standardized database of vector-based geologic map information that users can display and query via the Web, and then download selected data. The database will be a computer system distributed among all participating geological surveys in the United States. This technically difficult task has been proceeding slowly but steadily, as noted in various progress reports (Soller and others, 2001; Soller and Berg, 2002). The AASG and USGS are working together to build the foundation for the online map database, by developing the necessary technology, science concepts and data model, and by building digital geologic map coverage for the nation.

The NGMDB's Map Catalog and vector-based map database can be considered as endpoints on a spectrum of information content, ease-of-use, and system complexity. The Map Catalog is relatively straightforward to use, and since 1997 it has provided bibliographic information about each geoscience map product. In contrast, the online map database is in the design phase; when publicly available, it will provide full access to detailed geologic map information. However, many users may not be sufficiently familiar with geoscience concepts to comfortably use it's Web query interface. Further, many users will be unfamiliar with the format of GIS data to be offered for download; from user remarks regarding downloadable data now available from agency Web sites, we know this is frustrating.

We have the opportunity, somewhere in the middle of this spectrum, to provide users with geologic map information in a raster image format, thereby allowing them Web access to the familiar, traditional geologic map format (fig. 1). This is an important point, because the Web has increased the public's expectation for quick and reliable access to information. Although the Map Catalog informs the user that a map exists, and where it can be obtained, this is not sufficient for all purposes. Through discussions with users, and from comments received via our Web feedback form, it has become clear that many people are interested in viewing and/or obtaining maps online. Interpretation of the phrase "providing maps online" varies widely—to some people, it implies access to fully attributed, vector-based map databases, whereas to others, it implies access to map images. For example, a county planner contacted us because he could not view with common desktop office software the USGS topographic map

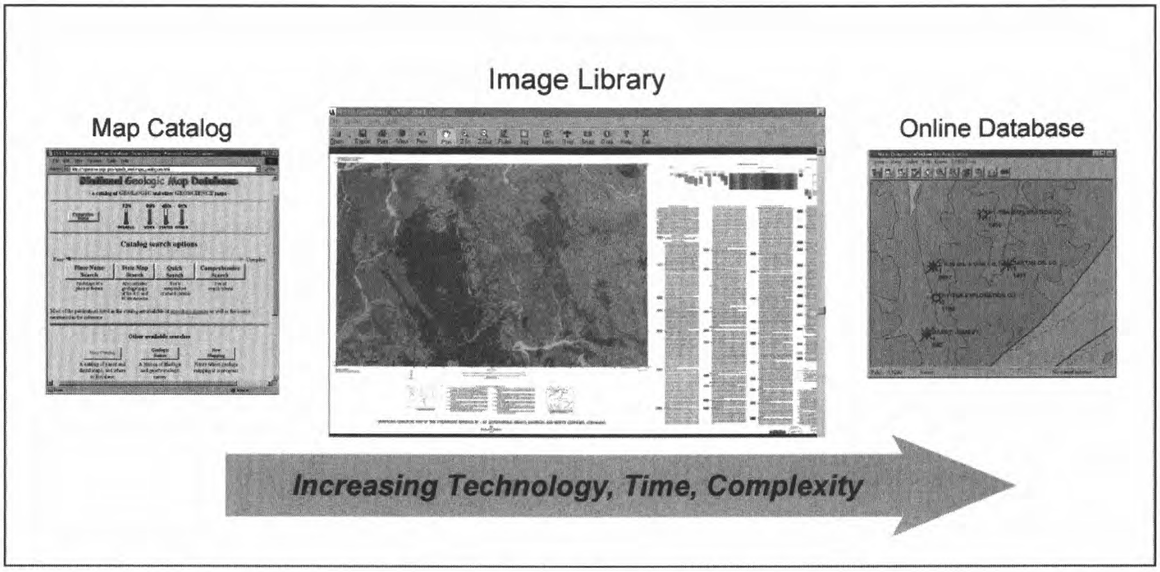


Figure 1. In technological complexity, ease-of-use, and related issues, the Image Library occupies a middle ground between the Map Catalog and the online map database.

data that he had downloaded. That file contained vector map data in the Spatial Data Transfer Standard (SDTS) format, which was not appropriate for the planner's needs—he simply wanted to view and browse an electronic facsimile of a paper map. This example highlights our mandate—to address the needs of all users, of various backgrounds and interests.

We are not yet ready to deploy a vector-based map database, but we can more quickly and efficiently develop the capability to let users view a digital image of a map online, and to read the detailed descriptions that accompany it. Therefore, the NGMDB has begun a new project initiative to build a library of images of general-purpose bedrock and surficial geologic maps. Map image libraries are not a new concept; they have successfully been developed by various agencies including the Library of Congress (Cahill and others, 2002) and state geological surveys (for example, Davidson and others, 2002). In this paper and in another paper in this volume (by Wardwell and others), we discuss the relation between the Image Library and the NGMDB Map Catalog, the capability for the Library to be a distributed database, and its technological foundation.

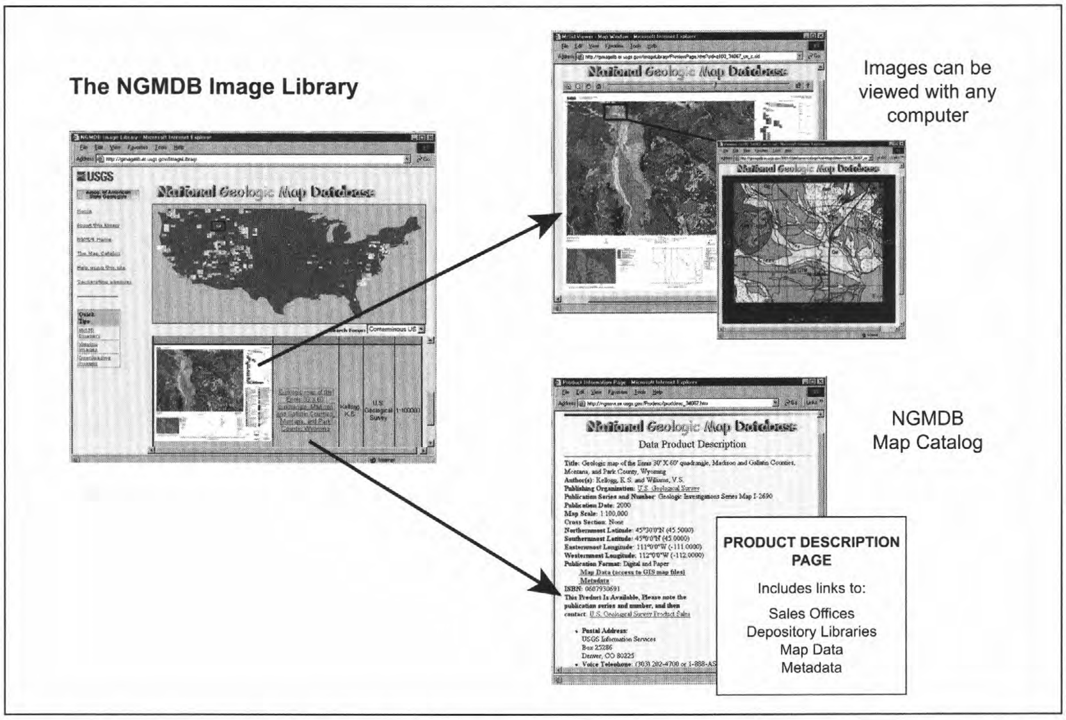
USER INTERFACE

The NGMDB's Image Library (see link at <http://ngmdb.usgs.gov/>) provides public access to images of selected general-purpose bedrock and surficial geologic maps, through an intuitive Web search interface. The search mechanism is geographically based, allowing for selection of geologic maps by location. From a base map of the U.S. or a state, the user can see where the NGMDB has map images available. When the user clicks on any of the geologic map thumbnail images displayed on the base map, the database is searched for all available scanned maps within a specified radius of the selected point (for example, within 25 miles). Results are returned to the user as an information table that shows a thumbnail image of each map within the search radius, and the map title, author, publisher, and scale (fig. 2).

If the thumbnail image is selected by a mouse click, a JPEG image is shown in a new window. Zoom and pan functions allow the user to read details on the map with excellent resolution, because each time the user zooms or pans, the high-resolution (300dpi) source image residing on the server is accessed and a 72dpi image, properly sized for the display window, is returned to the user (fig. 3). If we have permission to offer the map image for download at no cost to the user, a link in the new window offers this option (for example, all USGS maps are available for free download, but some agencies may need to charge a fee). The downloadable images are in the MrSID format, roughly 15-25MB in size; the 300dpi TIFF files from which the MrSID files were created are roughly 400-600MB in size, and are available to users upon special request.

If the user has a computer with the Windows operating system (Win98 or newer), the Lizardtech Express-View browser plugin enables viewing of the MrSID images, through the Image Library Web interface or by downloading the file and opening it locally in the Web browser. These options are provided by a link below the map thumbnail image. This browser plugin offers somewhat more efficient zoom and pan functions than the JPEG display, and the ability to zoom to (and past) the full resolution of the image.

In the information table, the map title is linked to its Product Description Page in the Map Catalog, which



Data Product Description

Title: Geologic map of the Ennis 30' X 60' quadrangle, Madison and Gallatin Counties, Montana, and Park County, Wyoming
Author(s): Kellogg, K.S. and Williams, V.S.
Publishing Organization: U.S. Geological Survey
Publication Series and Number: Geologic Investigations Series Map I-2690
Publication Date: 2000
Map Scale: 1:100,000
Cross Section: None
Northernmost Latitude: 45°30'0"N (45.5000)
Southernmost Latitude: 45°0'0"N (45.0000)
Easternmost Longitude: 111°0'0"W (-111.0000)
Westernmost Longitude: 112°0'0"W (-112.0000)
Publication Format: Digital and Paper
Map Data (access to GIS map files)
Metadata
ISBN: 0607930691
This Product Is Available, Please note the publication series and number, and then contact: U.S. Geological Survey Product Sales

- **Postal Address:**
  USGS Information Services
  Box 25286
  Denver, CO 80225
- **Voice Telephone:** (303) 202-4700 or 1-888-AS

Figure 2. The Image Library site. After selecting from the geographic search, an information table that shows all available map images is presented to the user. This table provides links to the JPEG and the ExpressView image-browsing interfaces (in this figure, both are shown – the more detailed map view appears in the JPEG interface). The table also links to the NGMDB Map Catalog product description page, map sales offices, a list of depository libraries, and downloadable map data and its metadata.

provides the user with bibliographic information and links to purchase the map, download it, or borrow it from a library. In the near future, we anticipate placing the thumbnail image of the map on the Product Description Page, and allowing links from the Map Catalog to the JPEG image in the Image Library.

STATUS

The NGMDB Image Library is being populated with MrSID-compressed images derived from 300dpi, 24-bit color scans that are stored off-line as TIFF files. We began the prototype Library by scanning 1:100,000-scale bedrock and surficial geologic maps published by the USGS. We now are scanning USGS maps published at 1:24,000-scale, and will include maps at other scales. At present, we are limiting our efforts to maps in paper format. However, we intend to serve (through a mechanism not yet determined) the images of maps that have been published only as digital datasets.

Because we have limited staff resources for this task, we have developed a database-management system (DBMS) to increase the efficiency with which both we and our cooperators can either load images into the Library, or link images to it. The DBMS is based on open-source architecture, and is described in another paper in this volume (Wardwell and others). In part, this system is intended to promote state geological survey participation, which is essential for the Image Library to become a comprehensive, useful resource. Contribution of maps to the Image Library is being designed as a simple process that involves Web forms and ftp data transfers. Agencies that would like to contribute to the Library should address two issues:

- will the map images reside on our (state agency) server, or on the USGS Image Library server?
- regarding user download of map images that are viewed through the Image Library, should we charge a fee for the map image, sell only the published paper map, or allow free download of the image?

If an agency wishes to charge a fee for downloading, or chooses not to allow downloading, this is quite reasonable and can be readily handled by the Image Library—as noted above, all maps in the Image Library are recorded in the Map Catalog, which provides direct links to agency map sales offices, so users can be advised of agency policies and fees.

If a cooperating agency chooses to manage the map images on their server, the Image Library DBMS would

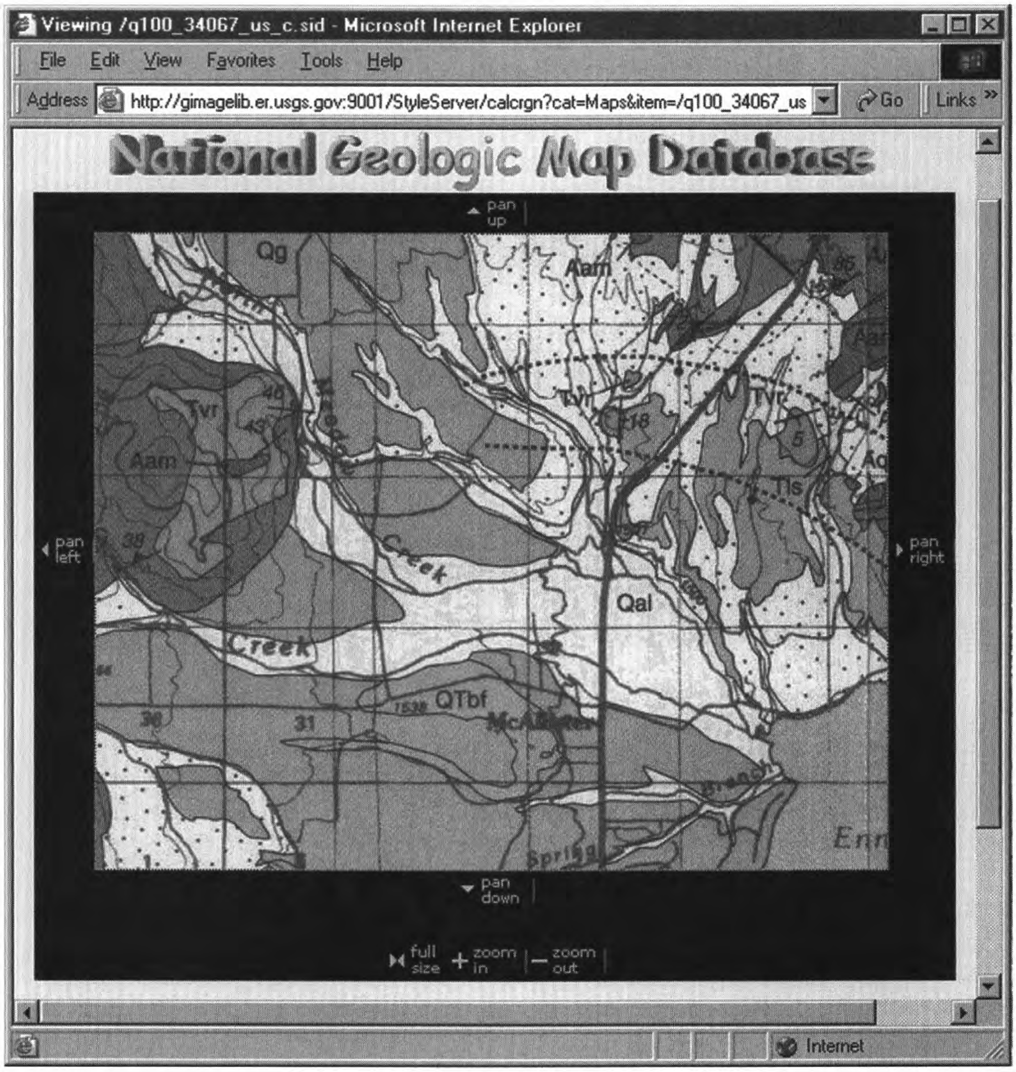


Figure 3. Example image from JPEG image-browsing interface, showing good-quality resolution. Image is from Kellogg and Williams (2000).

merely need to receive the URL and a thumbnail image for each map. To promote consistency in presentation of map images, we would prefer links to MrSID images. However, if the agency is charging a fee for access to the map, then they might choose to provide a URL only to a JPEG image of medium or low resolution.

If the agency prefers the USGS Image Library server to manage their map images, we would need only the TIFF and/or the MrSID images. If the agency wishes to participate but cannot create the scanned images, the USGS would scan selected agency maps and process them for the Image Library. Figure 4 shows these participation options.

SUMMARY

The USGS/AASG National Geologic Map Database (NGMDB) project has begun providing a Web-accessible Image Library of scanned geologic maps. This involves generating and processing scanned, high-resolution compressed imagery of selected general-purpose bedrock and surficial geologic maps that are recorded in the NGMDB Map Catalog. Scanned images of the selected geologic maps are formatted as compressed MrSID files, which can be viewed through a Web browser with no visible loss in clarity. The NGMDB Image Library is based on open-source architecture, and includes a database management system intended to promote participation by numerous agencies, principally the state geological surveys. The user interface is now available in prototype form, and we will continue to upgrade its map query options, user forms, and help pages. We hope this new NGMDB initiative further strengthens the cooperative relationship between the AASG and the USGS.

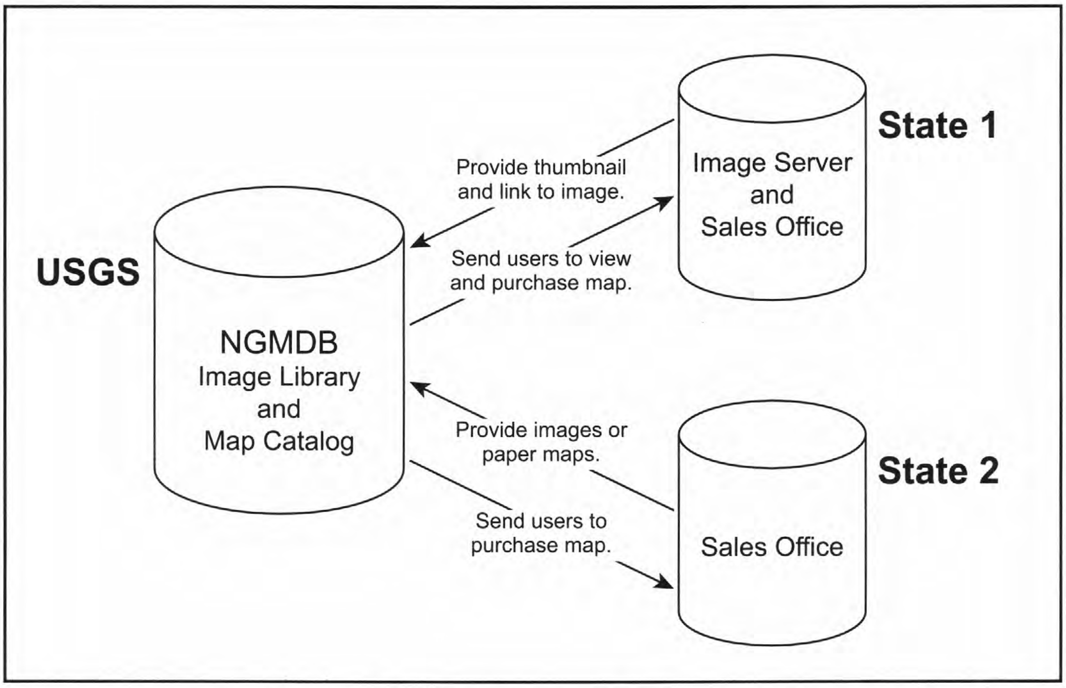


Figure 4. Participation by the USGS and state geological surveys is important in order to build a useful resource for public and private sector end users. State geological surveys choose how they wish to be involved.

REFERENCES

Cahill, C.R., Roberts, G.E., and Schug-O'Neill, Diane, 2002, Scanning and delivery of historic maps over the Web: The Library of Congress experience, *in* D.R. Soller, ed., Digital Mapping Techniques '02—Workshop Proceedings: U.S. Geological Survey Open-file Report 02-370, p. 99-104, <http://pubs.usgs.gov/of/2002/of02-370/cahill.html>.

Davidson, Gail, Staft, Lauren, and Daley, E.E., 2002, The Alaska DGGS scanning project: Conception, execution, and reality, *in* D.R. Soller, ed., Digital Mapping Techniques '02—Workshop Proceedings: U.S. Geological Survey Open-file Report 02-370, p. 105-110, <http://pubs.usgs.gov/of/2002/of02-370/davidson.html>.

Kellogg, K.S., and Williams, V.S., 2000, Geologic map of the Ennis 30' X 60' quadrangle, Madison and Gallatin Counties, Montana, and Park County, Wyoming: U.S. Geological Survey Geologic Investigations Series Map I-2690, scale 1:100,000, <http://pubs.usgs.gov/imap/i-2690/>.

Soller, D.R., and Berg, T.M., 2002, The National Geologic Map Database: A progress report, *in* D.R. Soller, ed., Digital Mapping Techniques '02—Workshop Proceedings: U.S. Geological Survey Open-file Report 02-370, p. 75-83, <http://pubs.usgs.gov/of/2002/of02-370/soller2.html>.

Soller, D.R., and Berg, T.M., 1999, Building the National Geologic Map Database: Progress and challenges, *in* Derksen, C.R.M., and Manson, C.J., eds., 1999, Accreting the continent's collections: Geoscience Information Society Proceedings, v. 29, p. 47-55, <http://ncgmp.usgs.gov/ngmdbproject/reports/gisproc98.html>.

Soller, D.R., Wahl, Ron, Weisenfluh, Jerry, Brodaric, Boyan, Hastings, Jordan, Laudati, Robert, Fredericks, Roger, 2001, Progress report on the National Geologic Map Database, Phase 3: An online database of map information, *in* D.R. Soller, ed., Digital Mapping Techniques '01—Workshop Proceedings: U.S. Geological Survey Open-file Report 01-223, p. 71-78, <http://pubs.usgs.gov/of/2001/of01-223/soller2.html>.

The National Geologic Map Database Image Library—Technical Details

By Robert S. Wardwell[1], Kevin W. Laurent[2], Jeremy O. Skog[3], and David R. Soller[3]

[1]U.S. Geological Survey
1300 SE Cardinal Ct.
Vancouver, WA 98683
Telephone: (360) 993-8908
Fax: (360) 993-8980
email: rwardwell@usgs.gov

[2]U.S. Geological Survey
Reston, VA 20192
email: klaurent@usgs.gov

[3]U.S. Geological Survey
Reston, VA 20192
email: drsoller@usgs.gov

INTRODUCTION

The Geologic Mapping Act of 1992 and its reauthorizations mandate creation of a National Geologic Map Database (NGMDB), to serve as a "national archive" of geoscience maps. The U.S. Geological Survey (USGS) and the state geological surveys (through the Association of American State Geologists, or AASG) are collaborating on the design and the many components of the NGMDB. Consistent with the Act's intention, the NGMDB project's overall objective is to provide the public with information and access to products that may be needed for purposes ranging from decision making to education and personal interest. The mandate is at least partially fulfilled through the NGMDB's Map Catalog (Soller and Berg, this volume; Soller and Berg, 1999) and its ancillary database, the Image Library. In this paper and in the companion paper in this volume, we discuss the relationship between the Image Library and the NGMDB Map Catalog, the ability for this library to link to outside resources, and its technological underpinnings.

In June, 2002, the NGMDB project began to build the Image Library, in order to meet the growing demand of Map Catalog users who simply wish to view a geologic map via their Web browser. The Image Library is now collecting and archiving scanned raster images of general purpose bedrock and surficial geologic maps. By archiving compressed digital raster images or links to available resources in various locations, the Image Library will provide a single vantage point from which users can find a geologic map to suit their needs. Once found, any map can be viewed through a standard Web browser as compressed image data in Multi-resolution Seamless Image Database (MrSid) format with no visible loss in image quality (see Soller and Berg article on the Image Library, this volume; fig. 3). Built-in zoom and pan capability allow the client to navigate a geologic map at multiple zoom levels without excessive pixelation. This allows users to clearly read even the finest details of the map, such as the legend descriptions and stratigraphic column information.

This paper discusses the technology behind the Image Library Web application, and the approach taken to meet the initial project requirements shown in figure 1. The Technical Overview section briefly describes the Library's computer infrastructure. For those readers who wish to learn more about the individual project components, such as database structure, Web content management, and online user interface development, please read the section on Web Application Development. Readers interested in learning more about how their organization can contribute maps to the Image Library will find this discussion in the companion paper in this volume.

TECHNICAL OVERVIEW

On a daily basis, the NGMDB Map Catalog is queried for records on bedrock and surficial geologic maps, and the query results are stored in the MySQL Image Library database. A paper or scanned digital copy of each

Image Library Project Requirements

1. Archive geologic map imagery by scanning paper maps and locating pre-existing maps in digital image format.

2. Store geologic map metadata from the NGMDB Map Catalog in a robust Image Library database.

3. Research the technology required to display geologic map image files over the Web.

4. Create a Web site allowing users to query imagery and map metadata from the Library and view maps, in detail, through a Web browser.

5. Within the Web site, develop a Library search tool allowing users to locate scanned imagery by referencing a map of the U.S. with multiple zoom levels. This map shows georeferenced thumbnail images of geologic maps stored in the Library.

Figure 1. Project requirements established early in the Image Library development phase.

map identified from the Map Catalog is then located. Most often, this is done by searching the library catalog at the USGS National Headquarters library in Reston, Virginia; maps are checked out of the library and scanned by USGS employees. In the future, the Library will include maps from a variety of sources, and provide an avenue for state geological surveys to disseminate their map imagery.

Paper maps are scanned full size at 300 dots per inch (dpi) and saved in TIFF format. These files are compressed at a ratio of 20:1 using the MrSID format. This technology was first developed at the Los Alamos National Laboratory (LANL) and later acquired by LizardTech Software, Inc. (http://www.lizardtech.com). MrSID imagery files for geologic maps are typically 15-25 MB in size. Both TIFF and MrSID format images are archived on CD/DVD media. One or several MrSID images for each publication are stored on an Image Library Web server so they can be accessed through a Web interface and viewed online.

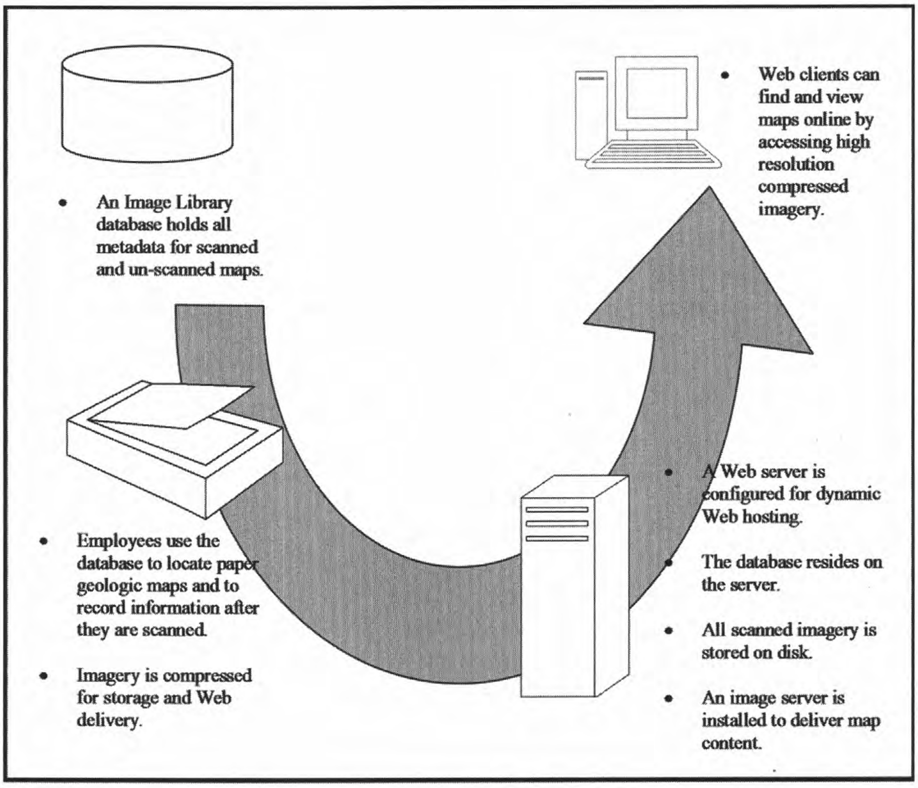
The Image Library currently contains 950 scanned bedrock and surficial geologic maps, with approximately 1300 raster images in MrSid format, accounting for publications that have multiple sheets. With over 23,000 USGS and state survey maps that may be scanned and put online listed in the database, it is apparent that the Image Library must be extensible and designed for growth. Careful thought has gone into designing a system that will allow a small staff to manage the large amount of raster imagery and metadata that must be collected, organized, and made easily accessible to the public via Web technology.

Figure 2 illustrates the major components of the Image Library IT infrastructure. The prototype Web server, located at the USGS National Headquarters in Reston, Virginia, hosts a relational database used to store limited metadata for each selected geologic map. The database also stores information about scanned imagery, and is crucial for tracking events and progress associated with map scanning. In order to grant users access to the contents of the database via the Web, the database has been linked to a Web content server and image server. This allows for newly acquired metadata and imagery to be dynamically incorporated into the Image Library Web pages. This system allows the staff to focus less on IT development and more on the acquisition of new imagery.

Figure 3 shows in more detail the hardware and software chosen to perform the actions described above. MySQL, a robust open source database, holds all map metadata (the map title, author, publisher, year, scale, geologic theme, media, and geographic bounding coordinates) and is regularly synchronized with the NGMDB Map Catalog Oracle database that is hosted in Flagstaff, AZ. The MySQL database holds links to a collection of MrSid format raster imagery residing on the server so that each image file is joined to the geologic map metadata.

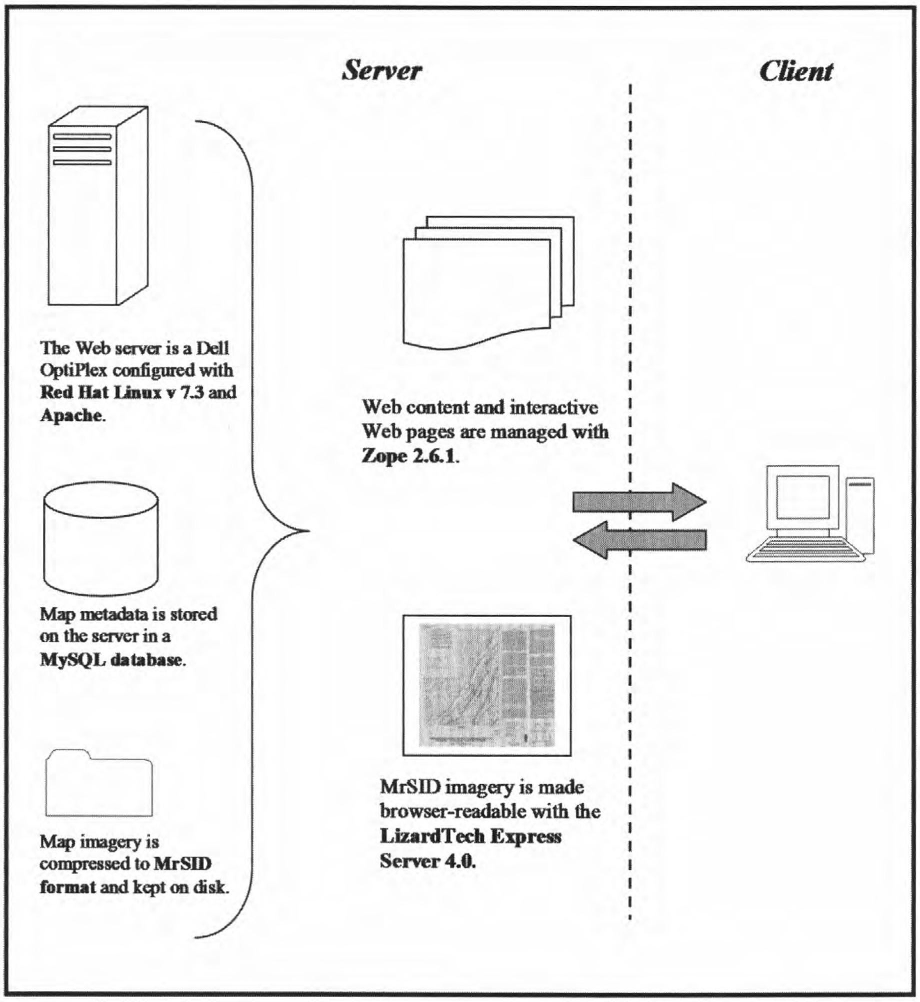
The LizardTech Express Server, a proprietary software for displaying MrSid images on the Web, is used to convert map imagery into a Web-friendly format (it delivers manageably-sized blocks of data to the user's Web browser) and to generate thumbnail jpeg images for each map. These thumbnail images, as discussed below, are important for maintaining the Web interface that enables users to search the Image Library.

Data from the MySQL database are brought together with the Web-ready imagery and transformed dynamically into Web pages by the Zope content management software. Zope is a powerful, open source solution for developing database-driven Web sites. With its own extensive object database, Zope allows Web program-



- An Image Library database holds all metadata for scanned and un-scanned maps.
- Employees use the database to locate paper geologic maps and to record information after they are scanned.
- Imagery is compressed for storage and Web delivery.
- A Web server is configured for dynamic Web hosting.
- The database resides on the server.
- All scanned imagery is stored on disk.
- An image server is installed to deliver map content.
- Web clients can find and view maps online by accessing high resolution compressed imagery.

Figure 2. Major components of the Image Library.



The Web server is a Dell OptiPlex configured with **Red Hat Linux v 7.3** and **Apache**.

Map metadata is stored on the server in a **MySQL database**.

Map imagery is compressed to **MrSID format** and kept on disk.

Web content and interactive Web pages are managed with **Zope 2.6.1**.

MrSID imagery is made browser-readable with the **LizardTech Express Server 4.0.**

Figure 3. Hardware and software chosen to manage the Image Library.

mers to insert bits of logic, such as a SQL database query, into a Web page so that users can interact with and search the database. In addition, Zope allows Website designers, those more concerned with the "look and feel" of the Website, the ability to create Web pages using standard HTML(Hyper Text Markup Language) code, or a WYSIWYG (What You See Is What You Get) editor.

Once user requests from the Web browser have been processed by Zope, standard HTML output is delivered to Web clients via Apache Web Server. The code needed to provide dynamic content is hidden to the user because Zope generates new HTML code each time a page is refreshed. Because Zope is so tightly integrated with the Image Library MySQL database, it also has been used to generate a Web accessible database front end (data input forms, query forms, and others) that the USGS staff uses to maintain metadata and imagery.

As seen in figure 4, a Web interface to the Image Library provides search tools that allow viewers to see *where* geologic map imagery is available. This geographic search interface is created without the help of a Web GIS server, such as ArcIMS, or a method for geo-referencing each individual map in the Library. Instead, the approach uses a unique combination of Extensible Markup Language (XML)-derived Scalable Vector Graphics (SVG) and JPEG image maps that are created directly from information in the database. This and several examples of how automated scripting has consolidated the work required to maintain the Web site are described in the following sections.

WEB APPLICATION DEVELOPMENT

Web Server

The first step towards developing the Image Library was choosing an adequate machine to act as a Web server, with enough disk space for a significant amount of data storage. A Dell OptiPlex GX400 PC with 160 gigabytes of disk space has been configured with the Red Hat Linux v.7.3 operating system. Red Hat Linux is a well-documented and tested operating system commonly used in Web server applications. Several versions of the operating system are freely available for download at <http://www.redhat.com>.

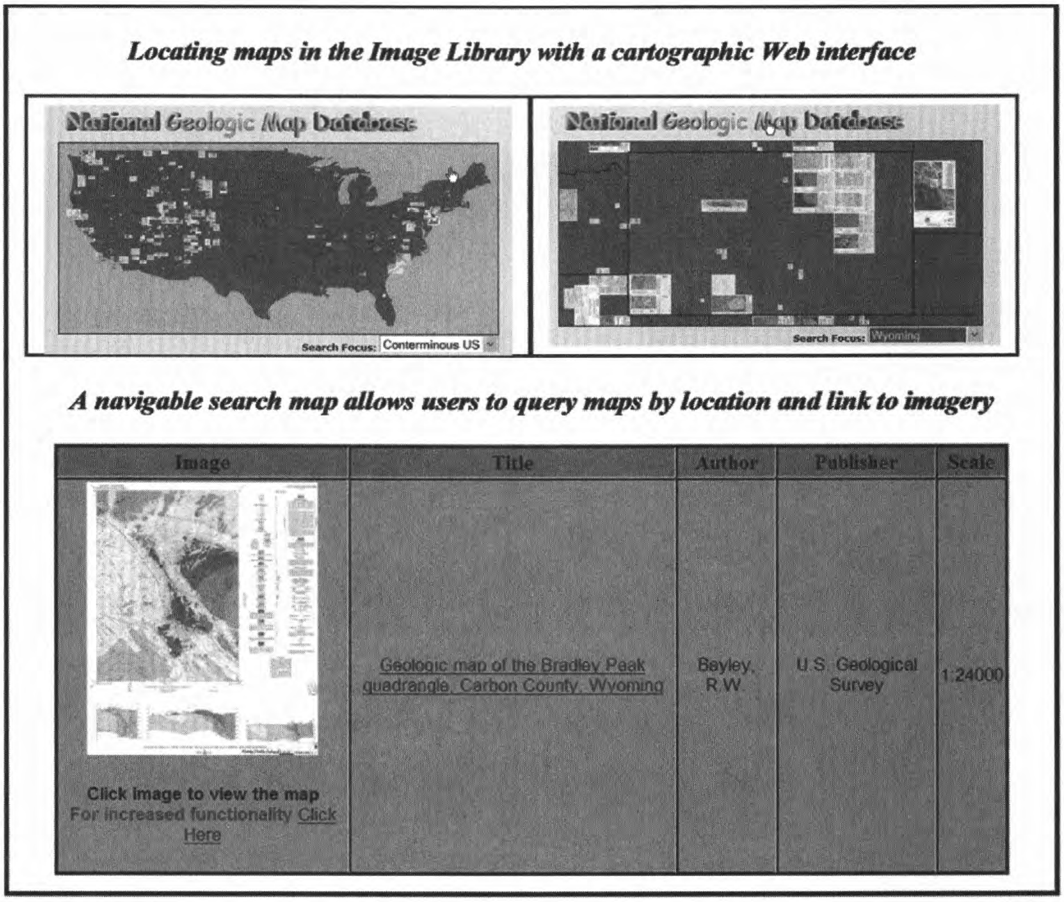


Image	Title	Author	Publisher	Scale
	Geologic map of the Bradley Peak quadrangle, Carbon County, Wyoming	Bayley, R.W.	U.S. Geological Survey	1:24000

Figure 4. The Image Library website is used to locate, view, and download geologic map imagery. An intuitive search map makes it easy to see where map imagery is available.

Database

MySQL v.4.0.1 is currently installed on the Image Library server to collect and store map metadata and to keep track of raster imagery. MySQL is the database of choice due to its ease of use and its track record in mission critical applications designed for the National Aeronautics and Space Administration (NASA) and companies such as Yahoo!, Finance, MP3.com, Motorola, Silicon Graphics, and Texas Instruments. The database is freely available for download at <http://www.mysql.com> and licensed through the GNU (GNU is a recursive acronym for "GNU's Not Unix") General Public License (GPL).

The MySQL Image Library database receives daily updates from the NGMDB Map Catalog Oracle 8.0.6 database through a series of shell scripts. These scripts, residing as chron jobs on the Image Library server, query the NGMDB Map Catalog for all geologic maps listed as "bedrock" or "surficial." Query output from the Map Catalog is formatted as tab delimited text and automatically loaded into the appropriate tables in the Image Library MySQL database.

Figure 5 shows how data is stored in MySQL. The database has been designed to perform three tasks. First it must store current information queried from the Map Catalog's Oracle database, as previously noted. This is handled by the *maps* table which contains a unique identifier, *pub_id*, for each map and its additional attributes. This table contains no duplicate records and stores the title, author, publisher, publication year, scale, media, bounding coordinates, and theme of each map. In the Map Catalog, each publication may have many themes (bedrock geology, surficial geology, coal, earthquakes). When scripts cause data to be loaded into MySQL, this theme information is processed so that each map is assigned an integer value key to the appropriate theme or themes in the *class* table. This allows the Image Library to label maps that are both "surficial" and "bedrock" in a single field.

The second task handled by the MySQL database is to track maps for which we have a scanned image, and provide a link to the image files on the server. The *scans* table is linked to the *maps* table on the *pub_id* field in a one-to-many relationship (because a map may have many map sheets). Because the MrSid imagery is stored in a server directory that is not part of the database, the *scans* table stores filenames that follow a standardized naming convention. This naming convention, illustrated in figure 6, is useful in managing imagery because the map scale, unique identifier (*pub_id*), sheet number, and publisher information can all be deduced from the filename.

In addition, the database helps the staff at the USGS keep track of maps as they are searched for in the library. If a map is not found in the USGS library card catalog, or has been checked out, this information can be stored using the *library* table so its status is known.

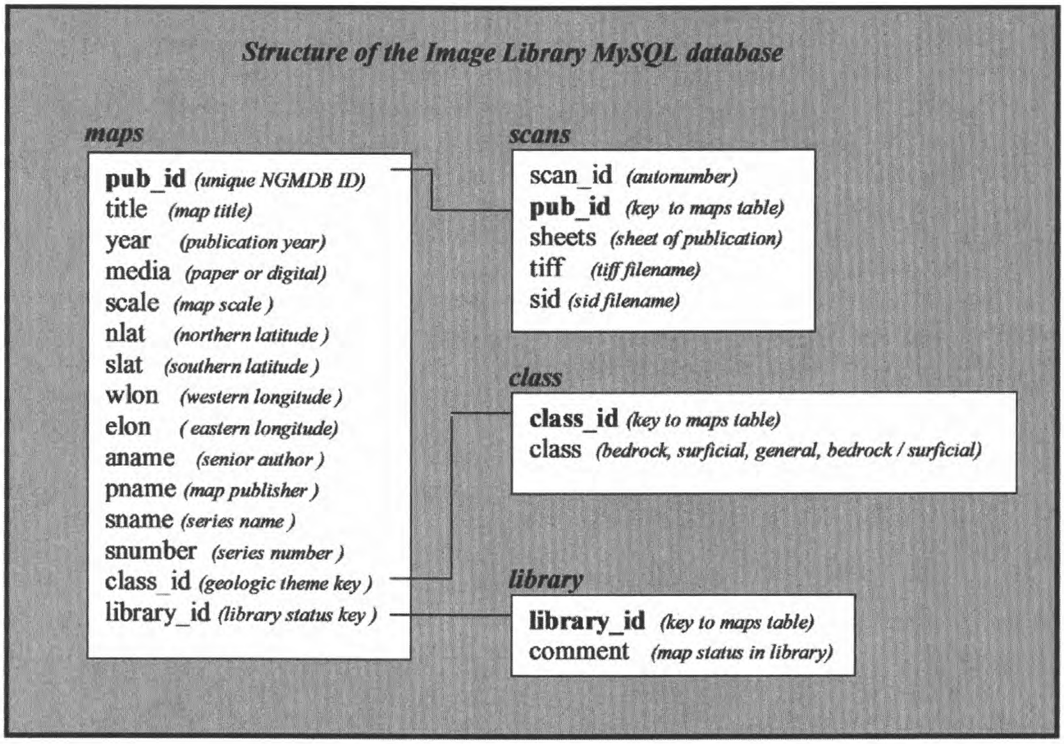


Figure 5. The MySQL Image Library database stores geologic map metadata, a listing of raster imagery, and a record of which maps have been scanned and which have not.

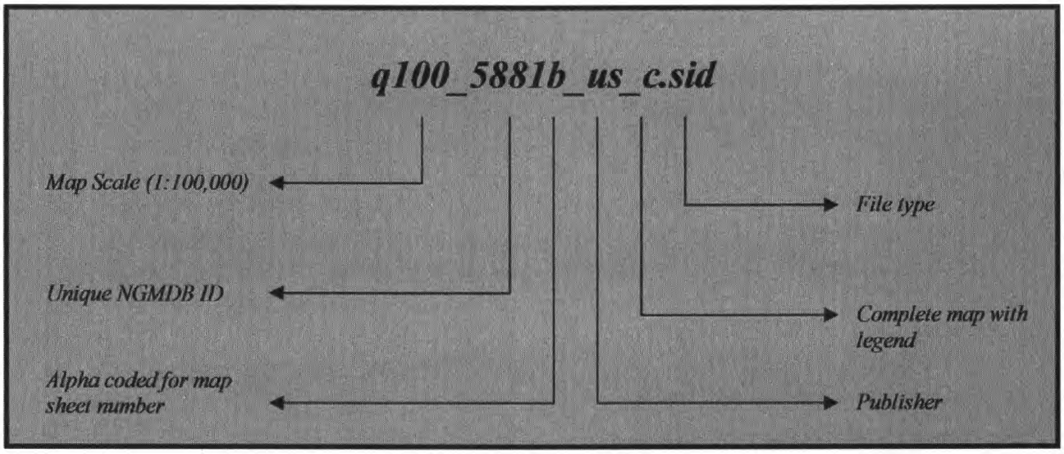


Figure 6. A standard file-naming convention is useful for keeping track of image files.

Image Server

As mentioned previously, when most geologic maps are scanned full size at 300dpi they can be stored in a Tag Image File Format (TIFF) file that typically ranges from 250 – 500 MB. By compressing these images into MrSID format, maps can be sized down to 15 – 25MB, but this is still much too large to be viewed over the Web. Even if the files were smaller, however, the MrSid is not a native browser format and a plug-in would be required to view the image online. In order to overcome these barriers, the Image Library employs the Lizardtech Express Server 4.0 which converts MrSID imagery in the Image Library to a browser friendly format.

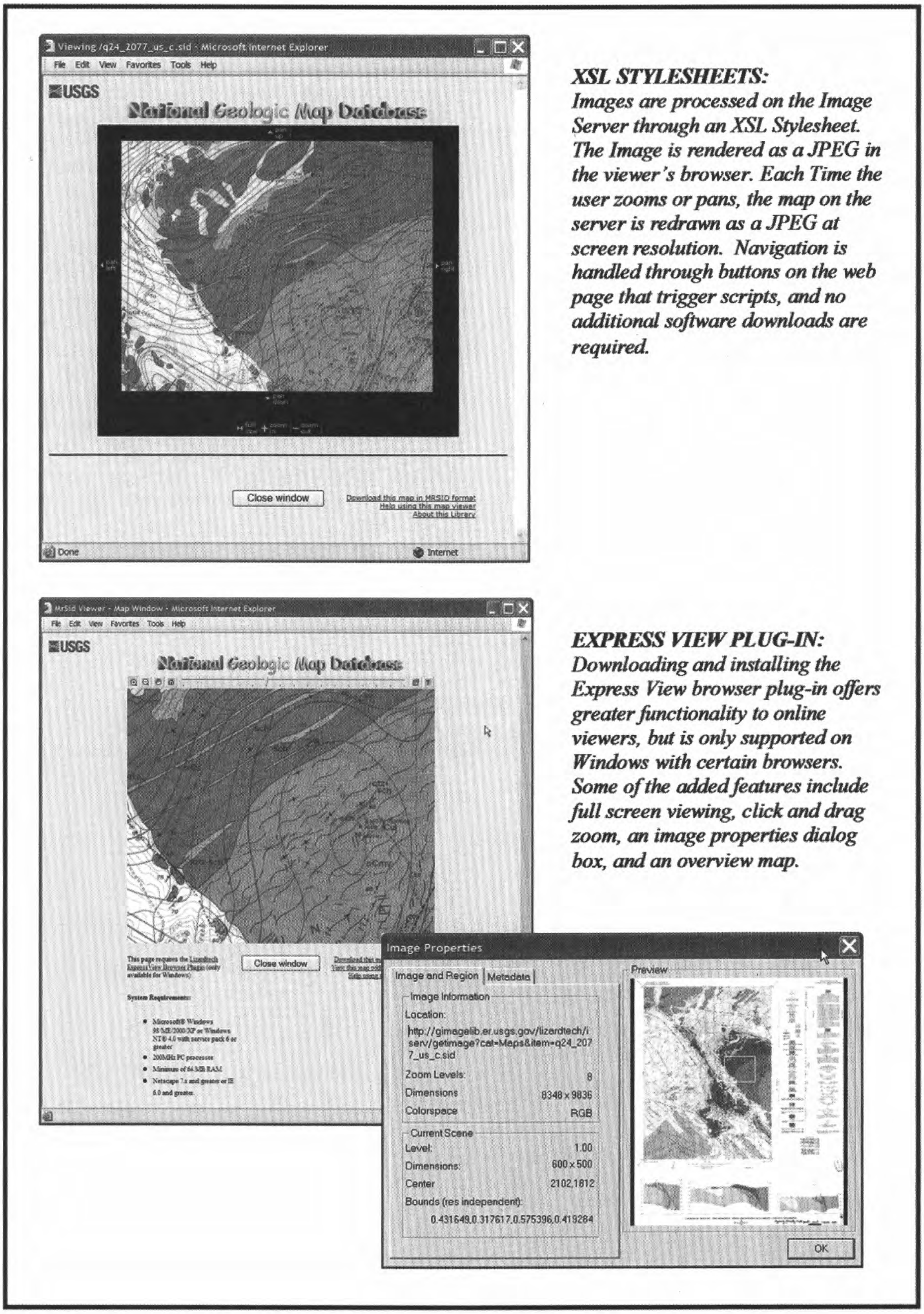
The Express Server, installed directly on the Image Library Web server, makes it possible for users to view maps at full resolution in one of two ways (fig. 7). The first method uses an Extensible Stylesheet Language (XSL) formatted style sheet and server side processing to dynamically render the MrSid image as a Joint Photographic Experts Group (JPEG) file in the viewer's browser window. The XSL stylesheet, which resembles a standard Web page in the user's browser, is configured by LizardTech, Inc. to operate with the Express Server package. Its "look and feel", however, is customized to the Image Library. Once rendered to a browser, the XSL stylesheet allows the user to zoom into and pan around the image to more clearly see the map and read the text. Each time the user changes their view of the map the image is redrawn. This method has proven to be quick and efficient in any browser with any connection speed.

The second method, available only to clients with Windows 98/NT/2000/XP, requires a download and installation of the Lizardtech Express View Plug-In. This software package enables MrSid format to be displayed through a Web browser and comes with several additional tools that enhance the viewing experience. For example, users can view the maps "full screen" or zoom into a map by performing a click and drag of the mouse. While plug-in performance is excellent over a fast Web connection it performs poorly over dial-up and is only available to Windows users. For this reason the Express View plug-in is only incorporated into the Image Library as an additional feature.

Web Content Management and Application Development with Zope

The *content* stored in the Image Library consists of a map metadata MySQL database and a collection of raster images in MrSID format. In a Web application sense, the ultimate goal of the Image Library is to allow users to ask for a map online (a *"request")*, and receive the image and metadata from the database (a *"response").* Often this Web / database interaction is handled by a server-side scripting language that processes the user input and generates dynamic Web output from the database. ASP, PHP, and JSP are just a few of the common server-side scripting languages found on the Web. Other available programs such as Zope and Cold Fusion have gone one step farther by allowing the Web application developer to manage all Web content, including the scripts that handle requests and responses, in single program environment. The Image Library uses such a program, Zope 2.6.1, to manage Web content and handle user / database interactions.

Zope is freely distributed by its designers, Digital Creations of Fredricksburg, MD (http://www.zope.org). The application features a transactional object database capable of storing content, custom data, dynamic HTML templates, scripts, relational database connectivity information, and code. For example, Zope's object database can hold an object named "sqlSelectMap", which is Structured Query Language (SQL) code that selects a map in the Image Library database from two input values: the latitude and longitude. Once the SQL query (called a *ZSQLMethod* in Zope) is stored, it can be inserted into a Web page anywhere with simple code and used to pass the results of the query back to any user who enters a latitude



XSL STYLESHEETS:
Images are processed on the Image Server through an XSL Stylesheet. The Image is rendered as a JPEG in the viewer's browser. Each Time the user zooms or pans, the map on the server is redrawn as a JPEG at screen resolution. Navigation is handled through buttons on the web page that trigger scripts, and no additional software downloads are required.

EXPRESS VIEW PLUG-IN:
Downloading and installing the Express View browser plug-in offers greater functionality to online viewers, but is only supported on Windows with certain browsers. Some of the added features include full screen viewing, click and drag zoom, an image properties dialog box, and an overview map.

This page requires the Lizardtech ExpressView Browser Plugin (only available for Windows).

System Requirements:

- Microsoft® Windows 98/ME/2000/XP or Windows NT® 4.0 with service pack 6 or greater
- 200MHz PC processor
- Minimum of 64 MB RAM
- Netscape 7.x and greater or IE 6.0 and greater.

Image Information
Location:
http://gimagelib.er.usgs.gov/lizardtech/iserv/getimage?cat=Maps&item=q24_2077_us_c.sid
Zoom Levels: 8
Dimensions 8348 x 9836
Colorspace RGB

Current Scene
Level: 1.00
Dimensions: 600 x 500
Center 2102,1812
Bounds (res independent):
0.431649,0.317617,0.575396,0.419284

Figure 7. The LizardTech Express Server can deliver MrSID imagery online by using XSL stylesheets or requiring a client side browser plug-in.

and longitude. Zope retrieves the query results, re-writes the HTML code with the new results, and sends the new Web page to the user. In a similar manner, Zope can store images, image properties, page design templates, database connections, scripts, text objects, etc. in a single environment where they can be rapidly accessed by the Web developers.

An important feature of Zope is that the development environment is accessed through a standard Web browser. This "through the Web" development model means that work can take place online from any Web-enabled computer on the internal USGS network. This activity is protected by an integrated security system that uses password encryption and a hidden directory structure. Multiple designers and programmers can access Zope's objects and make changes. Programmers can focus on bits of script and page functionality, while designers can edit the "look and feel" of the pages. In this sense, Zope allows the developers to separate "Web design" and "Web logic".

Figure 8 shows a glimpse of the Zope development environment, know as the Zope Management Interface (ZMI), viewed with Windows Internet Explorer. Each of the Zope objects, including folders, images, SQL queries, scripts, and page templates, are stored in a hierarchical manner as Web page links. New objects can easily be added using a pull down menu in the upper right hand corner. Clicking on these links allows the user to edit, view, test, set properties, and establish security constraints for each object.

Integrating the Zope objects together to build a finished Web application requires the use of two simple Zope scripting languages; Document Template Markup Language (DTML), and Template Attribute Language (TAL). Figure 9a illustrates how a developer might choose to manage Web content in Zope for a Website that allows online viewers to query a MySQL database. In this case the Zope objects are a database connection, a SQL query, an image with the NGMDB logo, and two HTML documents. Also included are two text objects labeled *standard_html_header* and *standard_html_footer* that hold the

Type Name	Size	Last Modified
☐ sqlSelectByMiles (select maps in search area)		
☐ sqlSelectByLoc_old (SQL for Map Selection by Location)		2004-01-02 15:19
☐ sqlSelectByLoc (SQL for Map Selection by Location)		2004-01-02 15:19
☐ sqlSelect50Miles (SQL for map selection within 50 miles of location)		2003-12-30 12:45
☐ sqlSelect100Miles (SQL for map selection within 100 miles of location)		2003-12-30 12:21
☐ mapfinder		2003-10-15 11:26
☐ getThumb_michigan		2003-12-05 14:46
☐ getThumb_hawaii		2003-10-02 18:47
☐ getThumb_Alaska		2003-10-27 19:25
☐ getThumb3		
☐ getThumb2		
☐ getRect (queries map BBOXs)		
☐ DatabaseConnect (Connect to the Im Library)		
☐ acl_users (User Folder)		
☐ table_bk.html		
☐ table2_.html		
☐ table2.html		
☐ table.html		
☐ selectResults (Results of Select by Lo		
☐ thinstripes.gif		

Title: select maps in search area

Connection Id: DatabaseConnect

Arguments:

```
target_lat:float
target_lon:float
search_rad_lat:float
search_rad_lon:float
```

```
SELECT distinct
m.pub_id,
m.aname,
m.pname,
m.title,
m.scale,
scans.sid,
scans.thumb,
c.class
FROM
maps m
LEFT JOIN class c ON (m.class_id=c.class_id)
LEFT JOIN scans ON (m.pub_id=scans.pub_id)
WHERE
(slat <= (<dtml-sqlvar target_lat type=float>+<dtml-sqlvar search_rad_lat type=f
(nlat >= (<dtml-sqlvar target_lat type=float>-<dtml-sqlvar search_rad_lat type=f
(wlon <= (<dtml-sqlvar target_lon type=float>+<dtml-sqlvar search_rad_lon type=f
(elon >= (<dtml-sqlvar target_lon type=float>-<dtml-sqlvar search_rad_lon type=f
(scans.sid <> "")AND
(m.pname In("U.S. Geological Survey","North Carolina Geological Survey"))
```

In the above view, a list of Zope objects in the Image Library folder is being accessed online via the ZMI. The first object in the list, sqlSelectByMiles, is clicked on to reveal the object as a standard SQL query in the window to the right.

Figure 8. The Zope Management Interface seen through Internet Explorer. The ZMI allows developers to quickly access and implement Web content.

bits of HTML code required at the top and bottom of each page on the Website.

In figure 9b, we see how these objects fit together using bits of code written in Zope's DTML language. The user is asked in *searchMaps.html* to enter a unique identifier, *pub_id*. This will return information about this publication from the database. When the user submits the entry, *yourMaps.html* is called. Within *yourMaps.html,* the SQL query *ZsqlSelectMaps* is run on the database, and the output results are written in standard HTML. Even though *yourMaps.html* only consists of several lines of DTML code in Zope it appears to the user as complete, standard, browser supported HTML code. When the next query is run, this HTML is modified by Zope and sent to the user once more. Also notice *NGMDBlogo.gif* appears automatically at the top of both pages because it is embedded in *standard_html_header.*

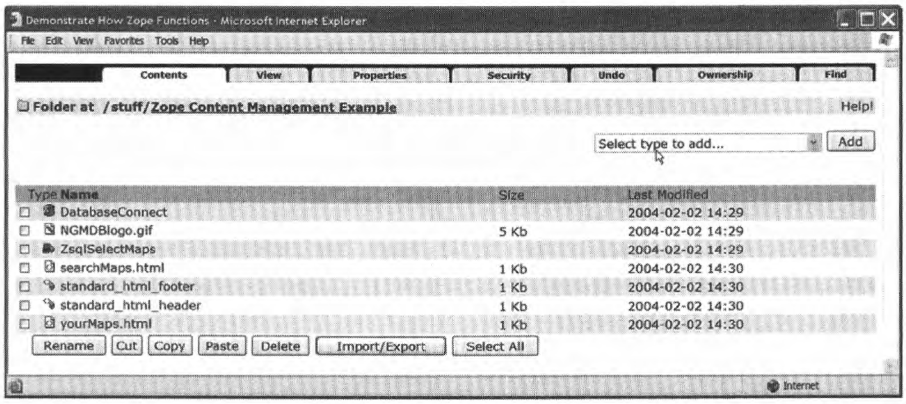
With its ability to manage client and database interactions through the use of stored objects, HTML input forms, and dynamically formatted HTML output, Zope has become a major component in the Image Library Web development. The Image Library staff also has implemented this technology to construct the fully Web-supported interface to the MySQL database shown in figure 10. This application is used to search the database for maps that have not been scanned, or to enter information about newly acquired imagery. It can be accessed from any location on the USGS internal network without the need for additional software.

One final item worth mentioning is that Zope includes its own relational database and Web server. For this project, however, Zope is configured with a more robust relational database, MySQL, and Apache Web Server supported by the LizardTech Image Server.

Online User Interface

The largest challenge of designing the Image Library has been developing an online user interface to meet the initial requirements of the project (fig. 1). The originally conceptualized interface (fig. 4) allows users to enter a Web GIS environment where they can see all available maps geo-referenced in their proper location. Adding multiple zoom levels allows users to get a closer look, and to eventually click on the geologic map they wish to see in detail. The imagery for the selected map, in MrSid format, is then loaded for online viewing. The main goal of this design is to allow for rapid access to scanned imagery by intuitive means.

While this design seems obtainable, and Web GIS applications are becoming more common all of the time, there are major obstacles in the development path. First, the Image Library eventually will host thousands of raster maps. The process of geo-referencing each map in the library adds a significant burden to the project workflow. Additionally, the process of loading, resizing, and rendering thousands of geo-referenced MrSid images into a Web GIS, such as ArcIMS, did not seem realistic. This opera-



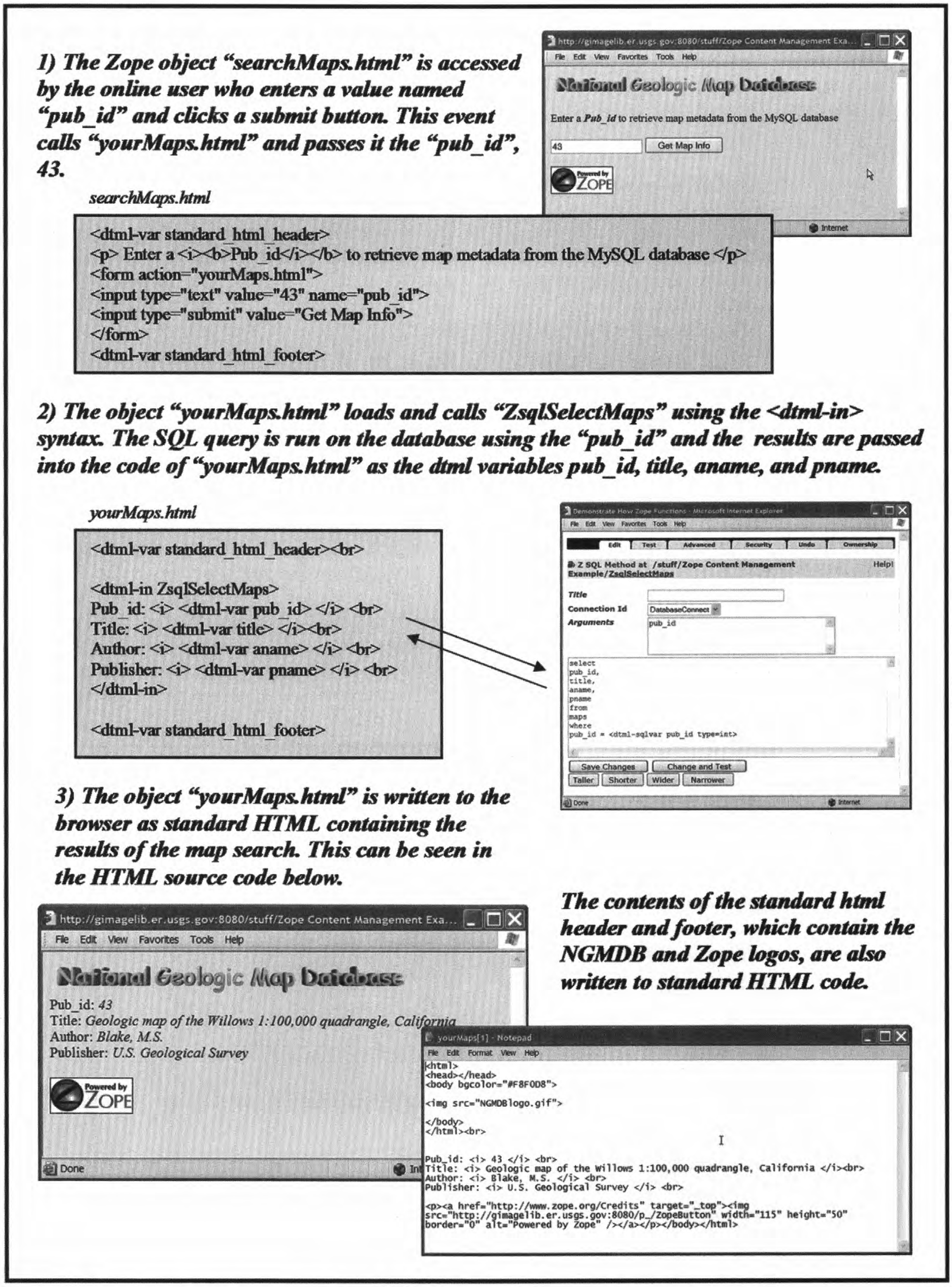
Folder at /stuff/Zope Content Management Example

Type	Name	Size	Last Modified
☐	DatabaseConnect		2004-02-02 14:29
☐	NGMDBlogo.gif	5 Kb	2004-02-02 14:29
☐	ZsqlSelectMaps		2004-02-02 14:29
☐	searchMaps.html	1 Kb	2004-02-02 14:30
☐	standard_html_footer	1 Kb	2004-02-02 14:30
☐	standard_html_header	1 Kb	2004-02-02 14:30
☐	yourMaps.html	1 Kb	2004-02-02 14:30

Through the ZMI we see seven Zope objects in a list. From top to bottom they are:

1. ***A connection to the Image Library MySQL database***
2. ***An NGMDB logo – (a GIF format image)***
3. ***A standard SQL query***
4. ***An HTML page used for searching the database***
5. ***An object that holds the standard HTML footer***
6. ***An object that holds the standard HTML header***
7. ***An HTML page used for displaying query results***

Figure 9a. An example of Web content storage in Zope.



1) The Zope object "searchMaps.html" is accessed by the online user who enters a value named "pub_id" and clicks a submit button. This event calls "yourMaps.html" and passes it the "pub_id", 43.

Enter a *Pub_id* to retrieve map metadata from the MySQL database

43

searchMaps.html

```
<dtml-var standard_html_header>
<p> Enter a <i><b>Pub_id</i></b> to retrieve map metadata from the MySQL database </p>
<form action="yourMaps.html">
<input type="text" value="43" name="pub_id">
<input type="submit" value="Get Map Info">
</form>
<dtml-var standard_html_footer>
```

2) The object "yourMaps.html" loads and calls "ZsqlSelectMaps" using the <dtml-in> syntax. The SQL query is run on the database using the "pub_id" and the results are passed into the code of "yourMaps.html" as the dtml variables pub_id, title, aname, and pname.

yourMaps.html

```
<dtml-var standard_html_header><br>

<dtml-in ZsqlSelectMaps>
Pub_id: <i> <dtml-var pub_id> </i> <br>
Title: <i> <dtml-var title> </i><br>
Author: <i> <dtml-var aname> </i> <br>
Publisher: <i> <dtml-var pname> </i> <br>
</dtml-in>

<dtml-var standard_html_footer>
```

Z SQL Method at /stuff/Zope Content Management Example/ZsqlSelectMaps

Title

Connection Id DatabaseConnect

Arguments pub_id

```
select
pub_id,
title,
aname,
pname
from
maps
where
pub_id = <dtml-sqlvar pub_id type=int>
```

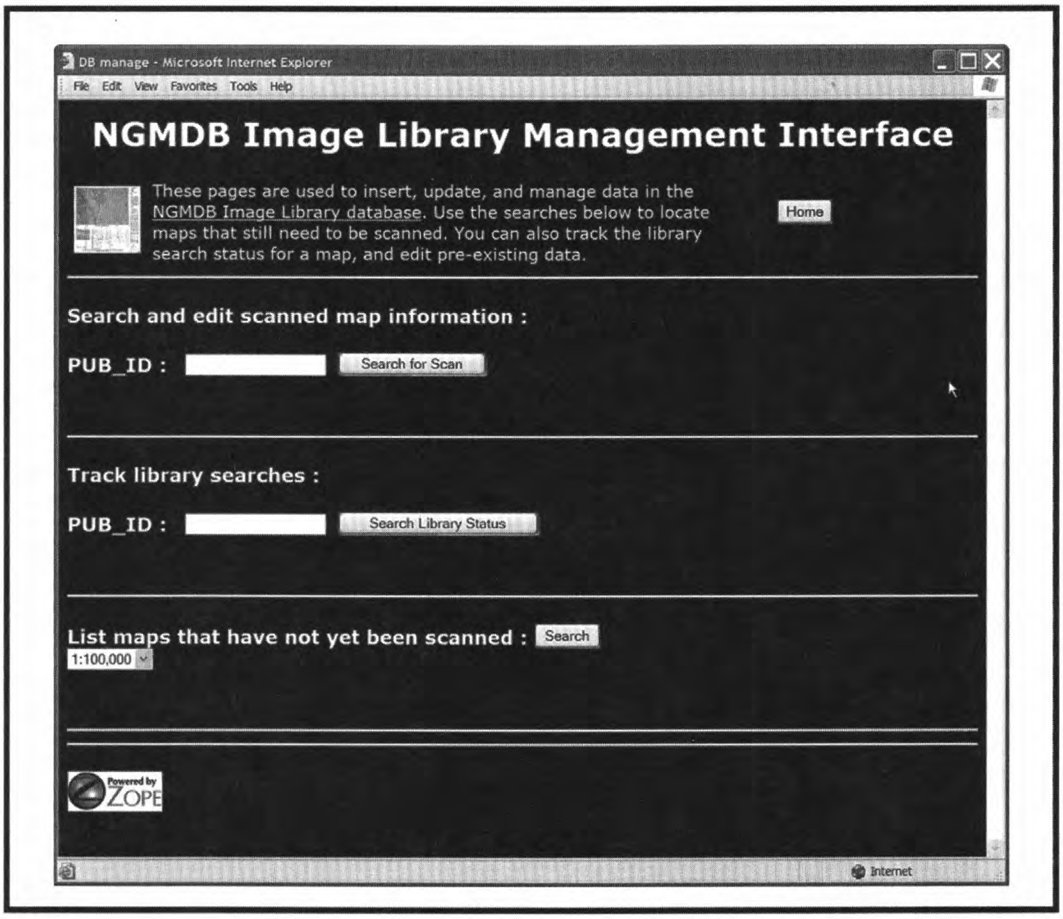
3) The object "yourMaps.html" is written to the browser as standard HTML containing the results of the map search. This can be seen in the HTML source code below.

Pub_id: *43*
Title: *Geologic map of the Willows 1:100,000 quadrangle, California*
Author: *Blake, M.S.*
Publisher: *U.S. Geological Survey*

The contents of the standard html header and footer, which contain the NGMDB and Zope logos, are also written to standard HTML code.

```
<html>
<head></head>
<body bgcolor="#F8F0D8">

<img src="NGMDBlogo.gif">

</body>
</html><br>


Pub_id: <i> 43 </i> <br>
Title: <i> Geologic map of the Willows 1:100,000 quadrangle, California </i><br>
Author: <i> Blake, M.S. </i> <br>
Publisher: <i> U.S. Geological Survey </i> <br>

<p><a href="http://www.zope.org/Credits" target="_top"><img
src="http://gimagelib.er.usgs.gov:8080/p_/ZopeButton" width="115" height="50"
border="0" alt="Powered by Zope" /></a></p></body></html>
```

Figure 9b. Dynamic Web pages generated with Zope that are used to search the MySQL database and return results to the user.



NGMDB Image Library Management Interface

These pages are used to insert, update, and manage data in the NGMDB Image Library database. Use the searches below to locate maps that still need to be scanned. You can also track the library search status for a map, and edit pre-existing data.

Search and edit scanned map information :

PUB_ID :

Track library searches :

PUB_ID :

List maps that have not yet been scanned :

Figure 10. The MySQL database management application used by the USGS staff to retrieve and enter data. The application is created with Zope and is accessed online over the USGS intranet.

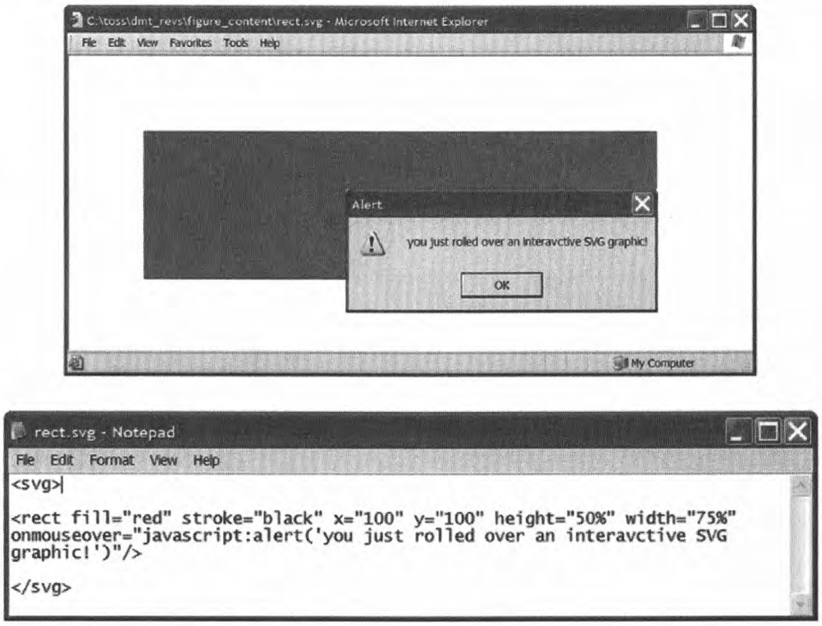
tion is time and resource intensive even in most *desktop* GIS applications. Other methods, such as generating HTML image maps using ArcGIS extensions and Arc Macro Language (AML) scripts seem possible, but once again, this approach requires a great deal of human interaction and maintenance.

As the project developed, the staff turned towards the use of Scalable Vector Graphics (SVG) for a solution. SVG, an open source XML-derived graphics format recommended as a future World Wide Web consortium (W3c) standard, uses XML grammar to draw pictures. In other words, images are created as a tag based text file resembling XML or HTML. It is useful for creating scriptable, animated graphical Web interfaces that resembles those made with Macromedia Flash technology. Because SVG graphics are text generated, they can be altered to reflect the contents of a database, and can be scripted to take full advantage of the Document Object Model (DOM). This aids programming of Web objects in HTML and JavaScript. Figure 11 shows a snippet of SVG code that draws a colored rectangle. The JavaScript written *into* the SVG image produces an alert message when a mouse cursor rolls over the rectangle.

Raster images can be embedded and placed at any specified location in the SVG image and pixel coordinates can be *transformed* into real world latitude and longitude coordinates. With a freely available SVG plug-in (Adobe's SVG viewer is available at <http://www.adobe.com/svg>) this format can be viewed inside of a Web browser window. SVG plug-ins also include zoom and pan functionality. When a user zooms into an SVG image, it is redrawn to the appropriate resolution so that image quality is not degraded. Many online examples illustrate how effective SVG can be when used in cartographic applications on the Web (see <http://www.svg.org/wiki/ow.asp?WebMappingExamples> and <http://www.carto.net>).

SVG is an excellent technology for the Image Library. The Library contains everything required to build an SVG file that takes each scanned image in the Library and draws it in geographic coordinate space, based on the latitude and longitude values stored in the database (see fig. 4). The process of writing an SVG Web map can be automated by formatting MySQL query output inside of the SVG text file.

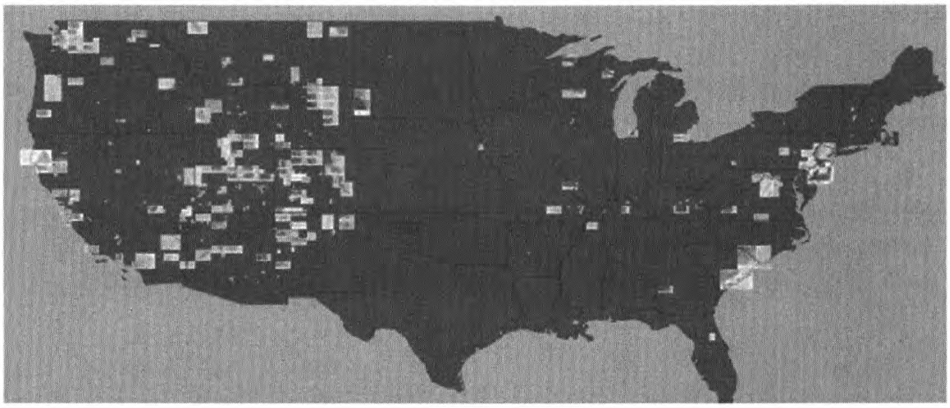
The map shown in figure 12 is an SVG representation of the Image Library database. It possesses full zoom



you just rolled over an interavctive SVG graphic!

```
<svg>

<rect fill="red" stroke="black" x="100" y="100" height="50%" width="75%"
onmouseover="javascript:alert('you just rolled over an interavctive SVG
graphic!')"/>

</svg>
```

Figure 11. A simple SVG image. This text-generated graphic is internally coded with JavaScript and displays a message when a rollover event occurs.



```
<g transform="translate(0,-89.0)">
<a xlink:href=http://gimagelib.er.usgs.gov:8080/ImageLibrary3/table.html?pub_id%3Astring=10000 target="table" >
<image xlink:href="http://gimagelib.er.usgs.gov/thumb/q100_10000_us_c.jpg
width="1.0"
height="0.5"
x="-107.0"
y="44.5"
name="10000"
border="0"/>
</a>
</g>
```

Each map stored in the SVG image is placed into a group tag (<g></g>) that can be shifted north or south with the "translate" attribute. As in HTML, the <a> tag denotes a link. Each small map image resides inside an <image> tag that links to a JPEG located on the server. Image attributes, such as x, y, width, and height are derived from values stored in the database so that maps can be placed in their proper geographic positions.

Figure 12. SVG is implemented to build a graphical representation of geologic maps available in the Image Library database.

and pan functionality and links each map in the SVG file to its raster image in the Library. Our original base map, showing the conterminous U.S. and state boundaries, is derived from an ArcGIS shapefile and saved in SVG / text format with Adobe Illustrator 9.0. A shell script stored on the Image Library server runs daily to produce a JPEG thumbnail of each MrSid file in the Image Library. The LizardTech Express Server makes this possible. Each JPEG thumbnail, sized at 256 pixels in the largest dimension, can be appropriately resized and positioned in the SVG image based on the bounding coordinates for each map stored in the MySQL database.

The code in figure 12 shows a section of the SVG source. The path for each map thumbnail is stored within an SVG image element using the *xlink* attribute. This tag points to the location of the thumbnail image on the Web server, and assigns a width, height, and location in SVG coordinate space. The SVG coordinate space is defined in the *svg element (<svg></svg>)* in the beginning of the file (not shown).

SVG seems like the perfect solution, but there is a significant problem common to many applications on the Web—not everyone uses the same browser. Internet Explorer, Netscape Navigator, Mozilla, Safari, Opera, and each of the other browsers and their many versions handle SVG content differently. Some do not yet handle the format at all, and the available plug-ins will not work in every browser. For this reason the Image Library Website has not fully implemented a SVG interface, but SVG is still an important part of the interface design.

As discussed, a SVG image representation of the database can be automatically generated through shell scripts and SQL queries stored on the server. This SVG image can, in turn, be captured as a JPEG that will load easily and quickly into a HTML page. When this technique is used at multiple SVG zoom levels, a collection of JPEG images is produced. Each JPEG snapshot of the SVG image represents a zoom level, and these images can be used efficiently in any browser setting to help users search for maps.

The process of automatically rendering a database-generated SVG as a collection of JPEG images relies on an open source product called the *Batik SVG Toolkit.* Batik (http://xml.apache.org/batik/) is developed under the Apache project by the same group that created the popular Apache Web Server. The Batik project has developed a group of Java modules that specifically deal with SVG. One of these modules, the *SVG Rasterizer*, allows for command line SVG to raster conversion with a variety of options.

The entire process can therefore be automated and set up to run at a specified time on the server. To create high-resolution, browseable JPEG images for the Image Library, a SVG of the conterminous U.S. is drawn from data in the database using shell scripts. The updated SVG image is opened with the Batik SVG Rasterizer and converted to a JPEG. To represent various zoom levels, the SVG is programmatically redrawn to a set of coordinates stored in an array and captured again as a JPEG by the Batik Rasterizer. Figure 13 provides a summary of how the Batik SVG Rasterizer is used to create browser supported

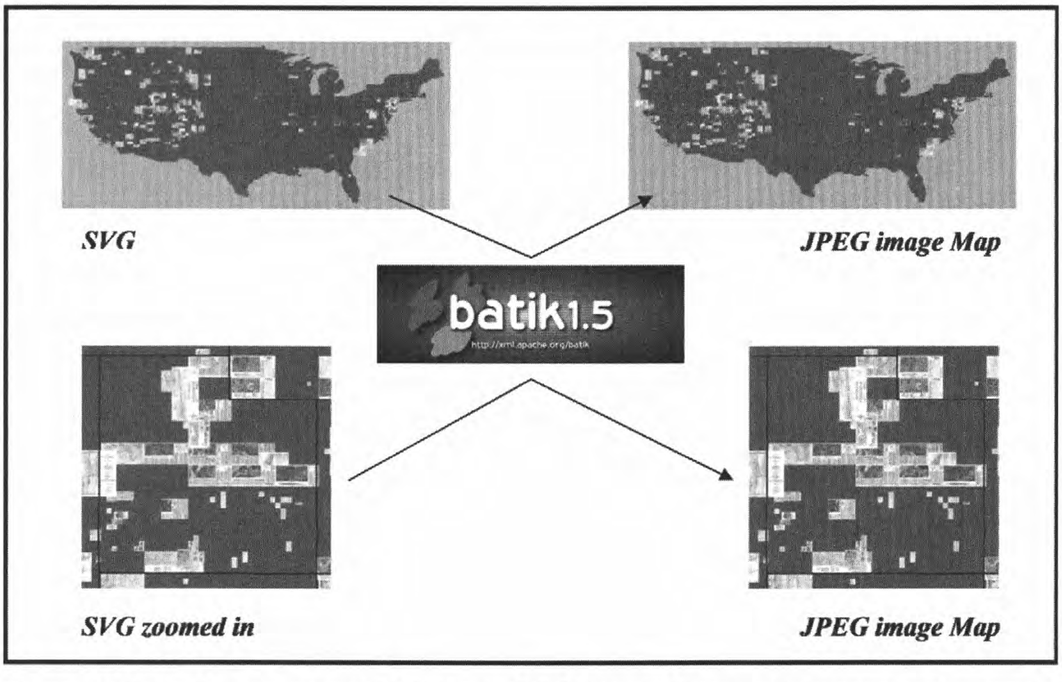


Figure 13. Using Batik to convert SVG files to JPEG format. When multiple zoom levels are processed, a collection of JPEG images is created. These images are used to build the Image Library cartographic search interface.

JPEG versions of an SVG image at multiple zoom levels.

Updated JPEG images are loaded onto the Image Library Web site as clickable image maps. This allows Web site viewers to specify an area on our searchable maps and retrieve a listing of available MrSID imagery from the database. Images at each zoom level are assigned a real world coordinate bounding box within a Javascript array. When a user clicks on the image map, the pixel coordinates are converted into the latitude and longitude coordinates used to search the database. SQL queries and page templates stored in Zope respond to the user's query and refresh the Web page with new query results. A link is provided to route users into a window where they can view the entire geologic map in MrSID format via the LizardTech Express Server.

We plan several new features for the Image library that will be added incrementally. First, a browser-compatible SVG interface will be added as technology allows, giving online Web users the ability to zoom and pan the search maps at greater magnification levels. An SVG implementation will also permit us to add data layers such as roads, populated places, and hydrology to help users locate maps more effectively. Several browsers are moving towards supporting SVG natively, as the technology becomes more popular.

SUMMARY

The National Geologic Map Database project is archiving raster imagery of geologic maps in MrSID format and making them available to the public via an online Image Library. The Image Library delivers complete geologic map information to the user's desktop where it is accessible as high resolution, lossless imagery. The main objective of the project is to encourage participation among USGS and state geological surveys. The initial Image Library Web release is scheduled for late 2003 or early 2004. In order to build a stable and robust system capable of future growth, the Image Library employs a MySQL database, a powerful Image Server from LizardTech, Inc. and Zope, a Web content management system. A Web site has been developed that allows users to see immediately on a national search map what maps are viewable, and to select the maps they wish to view by clicking on a location.

The Image Library is easy to use, and allows people to view geologic maps in great detail via the Internet. All USGS maps are available for free download. State geological surveys are encouraged to utilize the Image Library as a resource for their own maps, and have several options for participation (see companion paper in this volume).

This paper focuses on the technical details of how the Image Library Web application has been designed to increase efficiency and reduce the amount of human interaction and maintenance. From the beginning, our goal has been to move away from Web site development and into data acquisition and content. To support data acquisition, for example, we are working on a Web-based data submission form so that state geological surveys and participants within the USGS can submit images to the Image Library. The Image Library will then become a resource for organizations that wish to make their maps available via the Internet, but cannot justify the costs and labor associated with developing their own in-house systems.

REFERENCES

Cagle, Kurt, 2002, SVG Programming: The Graphical Web, Apress Publishing, p.1-20.

Cahill, C.R., Roberts, G.E., and Schug-O'Neill, Diane, 2002, Scanning and delivery of historic maps over the Web: The Library of Congress experience, *in* Soller, D.R., ed., Digital Mapping Techniques '02—Workshop Proceedings: U.S. Geological Survey Open File Report 02-370, p. 106-111. <http://pubs.usgs.gov/of/2002/of02-370/cahill.html>.

Lattier, Amos and Pelletier, Michael, 2000, The Zope Book: New Riders Publishing, p. 1-8. <http://www.zope.org/Documentation/Books/ZopeBook/>.

Lizardtech, Inc., 2003, Lizard Tech Express Server 4.0 FAQ: Lizardtech, Inc. company Web site. <http://www.lizardtech.com/support/faq/contentserver.php>.

MySQL AB, 1997-2003, MySQL Reference Manual, MySQL AB, p.1-19. <http://www.mysql.com/doc/en/index.html>.

Soller, D.R. and Berg, T.M., 2002, The National Geologic Map Database: A progress report, *in* Soller, D.R., ed., Digital Mapping Techniques '02—Workshop proceedings: U.S. Geological Survey Open File Report 02-370, p. 82-91, <http://pubs.usgs.gov/of/2002/of02-370/soller2.html>.

Science Language for Geologic-map Databases in North America: a Progress Report

By North American Geologic-map Data Model
Science-Language Technical Team[1]

ABSTRACT

A standardized language to classify and describe earth materials and their genesis is needed because producers and consumers of geoscience information use names, terms, and icons to communicate information about geologic objects and concepts. To the extent possible in a world of words, standardized terminology is useful to facilitate information exchange.

The Science Language Technical Team of the North American Data Model Steering Committee is a multi-constituency group of geologic-map producers and geologic-map users that, during the period April 2000 to November, 2003, has developed a prototype science-language for the naming and describing of earth materials in geologic-map databases produced by public-sector entities in North America. The classification adopts the following high-level architecture:

- Earth material
  - Igneous material
    - Hypabyssal rock
    - Plutonic rock
      - Plutonic
      - Hypabyssal
    - Volcanic rock
      - Intrusive
      - Effusive
      - Volcaniclastic
  - Composite-genesis material
    - Metamorphic rock (traditional sense)
    - Cataclastic rock
    - Impact-metamorphic material
  - Sedimentary material
    - Unconsolidated sedimentary material
    - Consolidated sedimentary material

These high-level categories fundamentally are genetic: they reflect how the earth material was formed (genetic process, crustal depth). This raises the irony that, although deeper levels of the classification hierarchies are based mainly on what the mapping geologist can see in the outcrop, traditional high-level classification approaches are compound, and link genesis with empirical composition and texture. Lower-level material names descending from each high-level category generally are based on singular textural or compositional criteria, depending on the parental category.

The use of standardized science language in digital geologic-map databases is a new frontier that is likely to evolve with time and experience. With this in mind, we are developing classifications of earth materials that we believe reflect not only how mapping geologists view them but also how such materials might be queried and analyzed in geologic-map databases. No single classification of earth materials will please all workers. However, the schemes we propose hopefully will be clearly understandable, internally consistent, and usable by both data-producer and data-user.

1 INTRODUCTION

1.1 Background

With the increasingly widespread production and use of digital geologic-map databases it has become clear that, to more effectively serve their constituencies, geoscience agencies need to develop several vital pieces of digital infrastructure:

(1) A standard conceptual model for storing digital data, and for manipulating these data in a relational and (or) object-oriented database environment;
(2) Standardized science language that allows geologic materials and geologic structures to be described, classified, and interpreted;
(3) Software tools for entering data into the standardized model at the front end (data-producer) and for extracting the data at the back end (data-user);
(4) Methodologies and techniques for exchanging data sets having different structures and formats.

To attain these objectives, public-sector geologic-mapping entities in the United States and Canada formed a partnership called the North American Data Model Steering Committee (NADMSC, http://geology.usgs.gov/dm/steering). This informal group is sponsored by cooperative agreements between the U.S. Geological Survey (USGS) and the Association of American State Geologists

---

[1] For information about this report, contact Jon Matti (jmatti@strider.swfo.arizona.edu).

(AASG), and between USGS and the Geological Survey of Canada (GSC). Through the former, NADMSC is linked to the database and standards-development activities of the National Geologic Map Database; through the latter, NADMSC is linked to database-development activities ongoing in Canada under the auspices of the Canadian Geoscience Knowledge Network.

The NADMSC first met early in 1999 to chart a strategy for developing various aspects of a standard geologic-map data model. Identified as a critical activity was the need for standardized science language for use in North America. To meet this task, NADMSC chartered a Science Language Technical Team (SLTT, http://geology.usgs.gov/dm/steering/science.html) that first convened in early 2000. SLTT members were identified in the following ways:

> (1) Most participants from the U.S. Geological Survey were identified by Regional Geologic Executives from the USGS Western, Central, and Eastern Regions. This group includes representatives of the geologic-map editorial standards units of the regional publications groups. Additionally, some USGS scientists were appointed by Coordinators of USGS line-item science programs;
> (2) Scientists from the Geological Survey of Canada and from selected Provincial geological surveys were identified by Canadian members of the NADMSC;
> (3) Scientists from State geological surveys were identified by the Digital Geologic-Mapping Committee of the Association of American State Geologists (AASG);
> (4) Scientists from the U.S. Forest Service, National Park Service, U.S. Bureau of Land Management, and Natural Resources Conservation Service were selected by the committee chair;
> (5) Academic members of the panel were selected by SLTT subcommittee co-chairs.

The assembled group (Table 1.1.1) represents a cross section of public-sector geologic-map producers and map users in the United States and Canada.

Table 1.1.1 NADMSC Science Language Technical Team committee members (Jonathan C. Matti, Committee chair)

Participant	Affiliation	SLTT Role
Lee Allison	Kansas Geological Survey	General scientific overview
Brian Berdusco	Ontario Geological Survey	General scientific overview
Thomas M. Berg	Ohio Geological Survey	General scientific overview
Sam Boggs, Jr.	University of Oregon	Sedimentary Subgroup
Eric Boisvert	Geological Survey of Canada	Sedimentary Subgroup
Andrée M. Bolduc	Geological Survey of Canada	Sedimentary Subgroup (co-chair)
Mark W. Bultman	U.S. Geological Survey	Sedimentary Subgroup
William F. Cannon	U.S. Geological Survey	Metamorphic Subgroup
Robert L. Christiansen	U.S. Geological Survey	Volcanic Subgroup (co-chair)
Jane Ciener	U.S. Geological Survey	Geologic-map editorial standards
Stephen P. Colman-Sadd	Geological Survey of Newfoundland and Labrador	Metamorphic Subgroup
Peter Davenport	Geological Survey of Canada	General scientific overview
Ron DiLabio	Geological Survey of Canada	Sedimentary Subgroup (co-chair)
Lucy E. Edwards	U.S. Geological Survey	Sedimentary Subgroup
Robert Fakundiny	New York State Geological Survey	General scientific overview
Kathleen M. Farrell	North Carolina Geological Survey	Sedimentary Subgroup
Claudia C. Faunt	U.S. Geological Survey	Volcanic and Sedimentary Subgroup
Mimi R. Garstang	Missouri Department of Natural Resources	General scientific overview
Joe Gregson	National Park Service	General scientific overview
Thomas D. Hoisch	Northern Arizona University	Metamorphic Subgroup
J. Wright Horton, Jr.	U.S. Geological Survey	Metamorphic Subgroup (co-chair)
David W. Houseknecht	U.S. Geological Survey	Sedimentary Subgroup
Bruce R. Johnson	U.S. Geological Survey	Volcanic and Metamorphic Subgroup
Robert Jordan	Delaware Geological Survey	General scientific overview
Ronald Kistler	U.S. Geological Survey	Plutonic Subgroup (co-chair)
Alison Klingbyle	Geological Survey of Canada	Geologic-map editorial standards
Dennis R. Kolata	Illinois Geological Survey	Sedimentary Subgroup
Elizabeth D. Koozmin	U.S. Geological Survey	Geologic-map editorial standards
Hannan LaGarry	Natural Resources Conservation Service	Sedimentary Subgroup
Diane E. Lane	U.S. Geological Survey	Geologic-map editorial standards

Victoria E. Langenheim	U.S. Geological Survey	Plutonic and Sedimentary Subgroups
Reed Lewis	Idaho Geological Survey	Plutonic and Volcanic Subgroups
Stephen D. Ludington	U.S. Geological Survey	Volcanic Subgroup (co-chair)
Jonathan C. Matti	U.S. Geological Survey	Sedimentary Subgroup (co-chair)
James McDonald	Ohio Geological Survey	Sedimentary Subgroup
David M. Miller	U.S. Geological Survey	Sedimentary Subgroup (co-chair)
Andrew Moore	Geological Survey of Canada	Sedimentary Subgroup
Douglas M. Morton	U.S. Geological Survey	Plutonic Subgroup (co-chair)
Patrick Mulvany	Missouri Department of Natural Resources	General scientific overview
Carolyn G. Olson	Natural Resources Conservation Service	Sedimentary Subgroup (co-chair)
Anne R. Poole	National Park Service	Plutonic and Sedimentary Subgroups
Stephen M. Richard	Arizona Geological Survey	Metamorphic Subgroup (co-chair)
Andrew H. Rorick	U.S. Forest Service	Sedimentary Subgroup
William Shilts	Illinois State Geological Survey	General scientific overview
David R. Soller	U.S. Geological Survey	Sedimentary Subgroup (co-chair)
Roy Sonenshein	U.S. Geological Survey	Sedimentary Subgroup
William C. Steinkampf	U.S. Geological Survey	Volcanic and Sedimentary Subgroups
Douglas B. Stoeser	U.S. Geological Survey	Plutonic Subgroup
Lambertus C. Struik	Geological Survey of Canada	General scientific overview
John Sutter	U.S. Geological Survey	General scientific overview
Harvey Thorleifson	Minnesota State Geological Survey	Sedimentary Subgroup
Robert J. Tracy	Virginia Polytechnic Institute and State University	Metamorphic Subgroup
David Wagner	California Geological Survey	Volcanic Subgroup
Richard B. Waitt	U.S. Geological Survey	Sedimentary Subgroup
Peter D. Warwick	U.S. Geological Survey	Sedimentary Subgroup
Richard Watson	U.S. Bureau of Land Management	General scientific overview
Gerald A Weisenfluh	Kentucky Geological Survey	Sedimentary Subgroup (co-chair)
Carl M. Wentworth	U.S. Geological Survey	Sedimentary Subgroup
Michael L. Williams	University of Massachusetts	Metamorphic Subgroup
Ric H. Wilson	U.S. Geological Survey	Volcanic and Plutonic Subgroup
Robert P. Wintsch	University of Indiana	Metamorphic Subgroup
Michael L. Zientek	U.S. Geological Survey	Plutonic and Metamorphic Subgroups

1.2 Related science-language efforts

SLTT activities benefited from a series of International Union of Geological Sciences (IUGS) sub-commissions chartered to develop uniform classifications of earth materials:

- *Igneous materials*: A long-standing IUGS Subcommission on the classification of plutonic and volcanic igneous rocks has led to a widely accepted standard (IUGS, 1973; MacDonald, 1974; Heiken and Wohletz, 1985; Schmid, 1981; Foley and others, 1987; Streckeisen, 1974, 1976, 1978, 1979; Le Bas and others, 1986; Le Maitre and others, 1989; Le Bas and Streckeisen, 1991; Le Maitre and others, 2002).
- *Metamorphic materials*: An IUGS Subcommission on the classification of metamorphic rocks (see http://www.bgs.ac.uk/SCMR/scmr_products.html) is underway, and is stimulating wide-ranging discussion of terminology for the naming, description, and genesis of metamorphic rocks.
- *Sedimentary materials*: An IUGS Subcommission on the geology of sedimentary materials (see http: //www.iugs.org/iugs/science/sci-cgsg.htm) is in the initial phases of its activities.

The International Union for Quaternary Research [INQUA] in the 1970's sponsored a Commission on Genesis and Lithology of Glacial Quaternary Deposits (Commission C-2). The results of Commission C-2 were published in Goldthwait and Matsch (1988; see Commission summaries in Goldthwaite and others, 1988, p. vii-ix, and Dreimanis, 1988, p. 19-25). The SLTT used this document to develop science language for sedimentary materials of glacial origin.

In a precedent-setting effort, in 1999 the British Geological Survey (BGS) issued four reports (Hallsworth and Knox, 1999; McMillan and Powell, 1999; Gillespie and Styles, 1999; Robertson, 1999) that presented science language for earth materials from a geologic-mapping point of view. The SLTT used these four reports as a starting point for our deliberations. The SLTT adopted major elements of the BGS approach, but found that in order to accommodate North American geologic-mapping traditions and approaches we had to develop slightly modified terminology and taxonomic hierarchies.

Within the United States, an important science-language activity is occurring under auspices of the Federal Geographic Data Committee (FGDC) Geologic Data Subcommittee (http://ncgmp.usgs.gov/fgdc_gds/). The FGDC has developed a draft cartographic standard for polygon, line, and point symbols that depict geologic features on geologic maps and digital displays. Although primarily concerned with cartographic technical specifications, the FGDC cartographic standard contains science-language concepts that ultimately must be integrated with and enlarged upon by the NADMSC SLTT group.

1.3 SLTT Housekeeping

SLTT has conducted its activities without dedicated salary and without a dedicated travel budget. As a result, face-to-face meetings generally were not possible, and SLTT members have boot-legged time from their agency science projects at the expense of project deliverables. The majority of SLTT interactions have been in the form of email discussions and conference calls. Both internal and external evaluation of science-language concepts was facilitated by a web-conference site that stimulated discussion of philosophical and operational issues (see http://geology.usgs.gov/dm/terms/).

Appendix 1 reprints the SLTT charter developed by NADMSC in 1999. Appendix 2 archives memoranda issued by the SLTT chair (J.C. Matti) discussing background issues and outlining guidelines for SLTT activities. These guidelines established the tone for Subgroup deliberations during the period April, 2000 through November, 2003.

Early on, SLTT decided to split into subgroups organized around major classes of earth material:

- Plutonic subgroup (R.L. Kistler and D.M. Morton, co-chairs)
- Volcanic subgroup (S.D. Ludington and R.L. Christiansen, co-chairs)
- Metamorphic subgroup (J.W. Horton and S.M. Richard, co-chairs)
- Sedimentary subgroup (J.C. Matti and G.A. Weisenfluh, co-chairs)
- Surficial-materials subgroup (A.M. Bolduc, R. DiLabio, D.M. Miller, C.G. Olson, and D.R. Soller, co-chairs).

Ultimately, SLTT recommended to NADMSC that the surficial and sedimentary subgroups merge into a single group, based on three factors: (1) unconsolidated surficial materials are sedimentary in origin; (2) the lithology, physical properties, genesis, and geomorphology of sedimentary and surficial materials are identical; and (3) scientific perspectives\and geologic-mapping experience in the two subgroups complemented each other and provided insights beneficial to both groups. NADMSC sanctioned this recommendation, and the combined sedimentary and surficial subgroups have worked together to develop a single body of science language for unconsolidated and consolidated sedimentary materials.

The SLTT chair selected subgroup co-chairs based on the following criteria: geologic-mapping experience, expertise in their science field, and knowledge of their agency's role in producing or using geologic-map databases. Subgroup co-chairs reflect a range of American and Canadian constituencies and Federal and State perspectives.

1.4 SLTT activities

20-queries exercise—SLTT's first order of business tasked each committee member to submit twenty queries to a hypothetical geologic-map database. This exercise had two purposes;

> (1) it served as a proxy for a requirements analysis that might be conducted among users of digital geologic-map data, to determine how such products are used and how the geologic data might be structured and organized from the point of view of content and language;
> (2) it was a means of getting each SLTT member to think about the science concepts that might be embraced by geologic-map databases, along with the issues and problems associated with naming, relating, and querying information about geologic materials and geologic structures.

Results of the 20-queries exercise revealed that database-users were interested in a broad range of geologic concepts and database targets—ranging from (1) academic queries related to the lithology, genesis, geometry, and age of geologic materials and structures to (2) pragmatic queries targeting what information geologic-map units and geologic structures contain about natural resources, fluid transmissivity (ground water and hydrocarbons), geologic hazards (swelling ground, landslides, earthquake-induced ground-shaking), and land-use planning (landfill siting, ground-water recharge, commercial and residential development, infrastructure siting). The SLTT's task was to develop science language to facilitate this broad range of potential database queries. Visit (http://geology.usgs.gov/dm/steering/teams/design/background.shtml) to examine the kinds of subjects reflected by the 20-queries experiment.

Results of the 20-queries exercise were passed along to the NADMSC Data Model Design Team (DMDT) for analysis and (especially) to ensure that science concepts emerging from the SLTT process were considered by

DMDT as it developed architecture for a standard geologic-map data model.

Iterative science-language development—Using the 20-queries exercise and building upon the four BGS classification documents, the SLTT subgroups iteratively developed science-language schemes that were exchanged by email among subgroup members. This process continued from about September, 2000 through March, 2003.

Internal SLTT review—After each subgroup completed a consensus classification of earth materials, subgroup documents were submitted for SLTT-wide peer review. This review was intended to ensure uniformity of philosophical and operational approach throughout the SLTT science-language process.

NADMSC review—Following internal SLTT-wide peer review, SLTT science language documents were forwarded to the NADMSC for evaluation and review for consistency, for geopolitical sensitivity, and for compatibility with the data-model architecture being developed by the DMDT.

Community-wide peer review—Following NADMSC approval, the SLTT documents are under final revision and will be released on a website for broad peer review from the North American geologic-mapping community.

1.5 Who prepared this report?

The SLTT chair (Matti) prepared this report in coordination with SLTT subgroup leaders (Table 1.5.1), each of whom contributed to the narratives in Section 3.

Table 1.5.1 SLTT Subgroup leaders who contributed to this report

Andrée M. Bolduc	Geological Survey of Canada	Sedimentary Subgroup
Robert L. Christiansen	U.S. Geological Survey	Volcanic Subgroup
Ron DiLabio	Geological Survey of Canada	Sedimentary Subgroup
J. Wright Horton, Jr.	U.S. Geological Survey	Metamorphic Subgroup
Ronald W. Kistler	U.S. Geological Survey	Plutonic Subgroup
Stephen D. Ludington	U.S. Geological Survey	Volcanic Subgroup
Jonathan C. Matti	U.S. Geological Survey	Sedimentary Subgroup
David M. Miller	U.S. Geological Survey	Sedimentary Subgroup
Douglas M. Morton	U.S. Geological Survey	Plutonic Subgroup
Carolyn G. Olson	Natural Resources Conservation Service	Sedimentary Subgroup
Stephen M. Richard	Arizona Geological Survey	Metamorphic Subgroup
David R. Soller	U.S. Geological Survey	Sedimentary Subgroup
Gerald A Weisenfluh	Kentucky Geological Survey	Sedimentary Subgroup

2 PHILOSOPHICAL AND OPERATIONAL APPROACH

2.1 Purpose

The SLTT purpose is to develop a science-language standard[2] for the description, classification, and interpretation of earth materials in geologic-map databases. The language should provide a logical, consistent, hierarchical framework for naming and classifying earth materials, and for describing their physical characteristics and genesis—based on the way geologic maps are made by the field geologist or assembled by a science compiler (Section 2.6.2).

2.2 Intended Use

Science language under development by the SLTT is intended for use by persons and agencies that submit digital geologic-map data into public-domain databases managed by various State/Provincial and Federal agencies. We are not setting standards for use by academia or by the private sector, unless these entities contribute geologic-map products to public databases.

Intended users include:

- geologists who collect original data in the field while making a geologic map;
- geologists who compile geologic-map data from

---

[2] *standard*—"*n.* **1**. Something considered by an authority or by general consent as a basis of comparison. **3**. a rule or principle that is used as a basis for judgment: *they tried to establish standards for a new philosophical approach.*
—*adj.* **23.** serving as a basis of weight, measure, value, comparison, or judgment. **24.** of recognized excellence or established authority. **25.** usual, common, or customary:..." (Webster's Encyclopedic Unabridged Dictionary of the English Language, 2001, p. 1857).

legacy sources and must interpret and translate these data for representation in the compilation;
- information-users who query public-domain geologic-map databases for information appropriate to their interests and applications.

2.3 Legacy data *versus* future data

The body of North American geologic-map information has two components: (1) "legacy data" archived in paper maps and digital files as the result of historic geologic mapping, and (2) new data that will be developed through future geologic-mapping. Incorporation of these two kinds of data sets into geologic-map databases involves different kinds of strategies, each posing its own challenge to science language.

North American legacy geologic maps are rich in geologic terminology. Typically, such data are contained either in map-marginal descriptions of map units or in pamphlets and reports that accompany the geologic map. Unfortunately, legacy maps rarely cite the classification systems used by the map maker to name and describe earth materials. Consequently, it is left to the map user to interpret the meaning and usage of terminology. For high-level terms (e.g., sedimentary rock, terrigenous-clastic sediment, plutonic rock, metamorphic rock, volcanic rock) the meaning may be universally understood. However, for deeper-level terms (e.g., shale, mud, basalt, quartz latite, quartz monzonite, granodiorite, volcaniclastic, slate, lahar, greenstone, gneiss, layered gneiss) the meaning may not be clear because many terms have inconsistent usage depending on when and where the map maker learned his or her craft. As a result, the map user commonly must interpret the meaning of earth-material terms according to his or her own experience.

This problem is compounded by two factors:

> (1) some geological terms have acquired usages that border on the generic or commonplace, and lack strict definitions or meanings (e.g., sandstone, granite, shale, gneiss);
> (2) some terms have been used as though they were lithologic names (e.g., alluvium, greenschist, till, turbidite, metasediment, loess, debris flow, lahar); this practice has blurred the distinction between lithologic description and genetic interpretation.

As a group, the SLTT committee had to wrestle with these issues, and decide on whether our science-language approach (1) should attempt to accommodate historical usage that is diverse, inconsistent, and in some cases generic, or (2) should reflect the needs and requirements of future geologic-map makers for science language that is stable and consistent. Ideally, any such decision will reflect the policy of the database developer, which usually means the management policy of the geologic-mapping agency or entity. With respect to legacy information, two contrasting data-management choices apply:

> (1) modern databases should archive and organize legacy terminology verbatim, without attempting to translate such terms into modern science language;
> (2) modern databases should interpret and translate legacy terms in the context of modern science-language structures, preserving archival terminology where it is clearly understood in terms of a modern standard but using more generalized terminology where the specific original meaning can not be reconstructed. NOTE: SLTT acknowledges that legacy geologic-map unit descriptions and supportive descriptions and interpretations must be archived *exactly* as indicated by the original map author. This should be accomplished by embedding legacy information in a text field or other field dedicated to such a purpose.

The SLTT group is not mandated to make such a policy decision on behalf of its constituent agencies. However, we recognize that legacy geologic maps include a wide variety of earth-material terms, many of which have similar, if not identical, meanings. Our purpose was to review how such terms have been used historically, and to judge how useful they are for storage, manipulation, retrieval, and analysis in geologic-map databases. In most instances, we found that traditional earth-material nomenclature lends itself well to database applications. However, we found that some traditional names and classification schemes did not adapt themselves easily to database requirements. In such instances, we had to modify existing names slightly, abandon some terms, or propose new names. The result is a hierarchical classification of earth materials that accommodates two objectives:

> (1) it allows legacy map terminology to be brought into modern geologic-map databases, using archival terminology where appropriate or by using generalized terminology where the specific original meaning is not clearly determinable;
> (2) it allows future geologic mappers to archive information about earth materials in a manner that is consistent, uniform, flexible, and forward-looking.

2.4 Operational Approach

The question of "What's in a name?" has plagued taxonomic classifications in all scientific arenas.

Historically, people have coined names for objects or concepts in order to convey information about them. The names are shorthand expressions (representations or proxies) for information packets that can be quite complex. Traditionally, the human brain has done the job of identifying all the attributes and components represented by a name. Now, we are asking computer databases to do this job.

The challenge to geologic classification and description in the database environment is:

- should a "name" express a range of concepts and attributes as it has historically? If so, then data-producer and data-user need to understand clearly that a specific name may represent a complex range of information content;
- should each "name" be relatively simple and explicit? If so, then other parts of the database structure must be used to store attributes that formerly might have been represented by a single name alone.

The prototype SLTT classifications lean toward the second approach. Each hierarchical level in the proposed classifications is designed to contain names or terms that represent geologic concepts that are as narrowly defined as possible. In general, we tried to avoid names or terms that represent complex combinations of geologic information. Instead, we strived to break these combinations down into individual attributes (descriptors) that are not used as part of the lithologic name but instead are relegated to other database fields.

This approach does not deny that most high-level terms, by their very nature, already are compound and complex. However, for deeper-level geologic names we tried to minimize their compound nature.

2.5 Definition of Concepts

The SLTT documents use certain concepts and terms (e.g., classify, name, define, describe) that have common generic meanings. For our purposes, these terms need to be delineated without ambiguity. The following definitions guided our deliberations:

Characterize—"v.t. **1.** to mark or distinguish as a characteristic; be a characteristic of....**2.** to describe the character or individual quality of....**3.** to attribute character to...." (Webster's Encyclopedic Unabridged Dictionary of the English Language, 2001, p. 347).

Controlled term—A term or name whose meaning and scope is restrained or restricted so that the term can be used or applied only according to the definition contained in a standard.

Classify—"v.t. **1.** to arrange or organize by classes; order according to class. **2.** to assign a classification to (information, a document, etc)" (Webster's Encyclopedic Unabridged Dictionary of the English Language, 2001, p. 381). To classify is to assign an *instance* to a specific group defined on the basis of a set of properties shared by members of the group. To *classify* answers the question "what kind of X is instance Y?", where X represents the domain of the classification.

Define—" v.t. **1.** to state or set forth the meaning of....**2.** to explain or identify the nature or essential qualities of....**3.** to fix or lay down definitely; specify distinctly....**4.** to determine or fix the boundaries or extent of....**5.** to make clear the outline or form of...." (Webster's Encyclopedic Unabridged Dictionary of the English Language, 2001, p. 523).

Description—A set of statements that characterize the nature of a thing (a class or instance) such that the thing may be identified and named.

Earth material—A naturally occurring substance formed in or on the Earth by physical, chemical, or biogenic processes that produce solid particles or crystals of mineral and (or) rock.

Instance—"n. **1.** a case or occurrence of anything...." (Webster's Encyclopedic Unabridged Dictionary of the English Language, 2001, p. 988).

Geologic-map unit—An intellectual construct that a geologist delineates on a map as a way to communicate a geologic concept to the map user. Each geologic-map unit corresponds to a three-dimensional volume of earth material that consists of one or more discrete lithotopes whose character and (or) frequency of occurrence makes each map unit distinct and unique from other such units. The map maker defines the scope, scale, boundaries, names, and reference sections for geologic-map units according to rules developed and adjudicated by the North American Commission on Stratigraphic Nomenclature (NACSN, 1983).

Lithotope—A body of sediment or rock that can be "a stratigraphic unit, a part of a stratigraphic section, a particular kind of sediment or rock, [or] a body of uniform sediments formed by the persistence of the depositional environment" (*Glossary of Geology*, Jackson, 1997, p. 373).

Modifier—A term or word that limits, constrains, or qualifies a *controlled term*.

Name—"n. **1.** a word or a combination of words by which a person, place, or thing, a body or class, or any object of thought is designated, called, or known...." (Webster's Encyclopedic Unabridged Dictionary of the English Language, 2001, p. 1276). By this definition, a name is a proxy for the complete description that defines the nature of something (e.g., an "instance" of a geologic map unit). Ideally, every instance would have a unique name, but of course in the realm of earth materials, this is not always the case. That is why SLTT attempts to provide a standardized classification of names for earth materials, their physical attributes, and their genesis.

2.6 Guidelines followed by SLTT

In developing science language for earth materials, SLTT adopted the following rules:

2.6.1 Descriptive classification basis

To produce a classification system for earth materials that allows different observers to classify a given sample in the same way, the system must be based on physical properties of the rock recognizable by all observers. The properties used for field classification of earth materials include modal mineralogy, grain size, grain shape, rock fabric (the arrangement of grains in an aggregate to form the rock), and structures in the rock (bedding, layering, etc.). Although distinct bodies of rock may be recognized based on other physical properties (e.g., magnetic susceptibility, or density), these generally are not used as field criteria.

The approach to a lithologic classification developed here is fundamentally descriptive—that is, classification of an earth material is based on its observable features, and its assignment to a lithologic class implies that certain descriptive criteria are met. These descriptive criteria provide default attributes for each earth material, and no other attributes are required to satisfy the material definition. For example, the name for a sedimentary material (e.g., sandstone, calcareous dolostone, slightly gravelly sand) is a proxy for a default description embedded in the database simply through application of name to the material.

2.6.2 Science language compatible with geologic-mapping strategies

The goals and methods of geologic mapping require science-language structures that are different from those of other endeavors. This is because the types of observation that go into developing a geologic-map database do not have the same information content and scientific credibility at all locations within the map footprint. This conclusion is based on the nature of the geologic-mapping process.

For each geologic map, the scope, scale, and consistency of geologic observation varies throughout the map footprint. This is because the nature and quality of each observation varies from place to place, depending on its purpose, the time available to make it, and the quality of the geologic outcrop. Many observations upon which the map is based are detailed and comprehensive; others are generalized and cursory. The latter typically are not the fault of the geologic-map maker, but rather are intrinsic to the geologic-mapping process itself: every potential observation point within the map footprint cannot be examined with the same level of definitive care and quality, and the information content within a geologic-map unit must be extrapolated between observation points—some of which may be quite far apart.

Consider the type of observation a mapping geologist might make in determining whether a particular outcrop should be included within a particular map unit or excluded from it:

- Binocular observation of a distant outcrop series to determine the ratio between 'sandstone' and 'conglomerate' ("Looks to me like 'sandstone' dominates over 'conglomerate');
- Casual observation of grain-size ratios in an outcrop in order to confirm that lithologic trends in a series of outcrops still apply ("Looks like the same old 'sandstone' beds. Don't need to examine these very carefully");
- Detailed hand-lens determination of grain-size ratios in a series of sedimentary beds in order to characterize a given outcrop in detail ("These 'sandstone' beds look a little different from the previous ones; I should spend some time and compare them to those in the preceding outcrops, just to be sure they belong in the same map unit");
- Follow-up petrographic analysis to determine details of texture, fabric, and grain mineralogy ("Even though I described these beds in the field as 'sandstone' based on hand-lens observation, I see on the basis of microscope observation that the mud-size fraction is greater than I originally believed. These beds more properly should be termed 'muddy sandstone', and they are more akin to the mudstones of Formation Y than they are to the sandstones of Formation X").

The preceding examples suggest that a hierarchical observational approach characterizes the geologic-mapping process—ranging from generalized observations that are reconnaissance in scope to detailed observations that are definitive in scope. Each observational style has its own confidence level. Moreover, science-language terms for each observational level have slightly different meanings depending on the scale of observation. In each of the preceding examples, (1) does the term "sandstone" have the same meaning? (2) Are different types of information communicated through the use of "sandstone" in each circumstance? We answer "no" to the first question and "yes" to the second. This is a different process than takes place in the controlled environment of a petrology laboratory, where specific kinds of questions are pursued systematically and answered using language that is definitive and precise. Accordingly, we conclude that the science language of earth materials in geologic-map databases must be structured to reflect the hierarchical nature of observations made during the geologic-mapping process.

This is not to say that the observational quality of geologic-map information is poor—in specified places it is quite good. However, developers and users of science language for geologic maps need to be aware that (1) not all observations have the same level of refinement and (2) information projected (extrapolated) outward away from observation points without benefit of intervening data—the essence of geologic-map making—is vulnerable. These limitations require the science language of geologic-map data sets to be constructed so that the language reflects how geologic mapping actually is executed.

2.6.3 Progressive Hierarchical Structure

The classification language should be progressive: that is, it should be based on what the geologist can observe and describe sequentially during the course of making a geologic map—first with the un-aided eye, then with hand lens, and then with thin section. Each of these observation classes yields a package of information that differs in scope, content, and rigor from that in other classes. Lithologic names should be developed that are consistent with each observational class.

The progressive nature of the observation process yields a hierarchical language structure—that is, language that begins at a generalized level and progressively has more specific categories that communicate more refined information about each earth material. This hierarchical structure must rigidly follow the rules of parent-child lineages—that is, each child should occur only once in the hierarchy, and should have only one parent. This is important because compound parentage (where a geologic object can have more than one parent—i.e., can be interpreted as the result of more than one process) makes the classification process difficult and can lead to misleading hits during database queries and analysis.

Developing a logical hierarchical structure proved to be vexing. As with Linnean zoological taxonomy, the purpose of organizing earth-material names into parent-child lineages is to identify logical relationships among individual lithologic types and groups of lithologic types; taxonomic names presumably should reflect these relationships. In the case of geologic-map databases, the premise is that lumping and splitting real-world objects into inter-related categories will help in analyzing the objects, and will facilitate searching the geologic-map data set for items as narrowly or broadly defined as our interests require. We assume this premise is a valuable one, and that a hierarchical classification approach is not just a clerical device but has functional utility.

2.6.4 Clarity and Ease of Use

The data-producer and the data-user must understand clearly the basis for the earth-material classification schemes, and must be able to use them easily and comfortably. In some cases, proposed SLTT language structures require the map maker and map user to re-think or re-learn how to use some terms. However, in general, we predict that our language structures will be familiar to most users, and will provide a systematic and uniform way to describe and interpret earth materials.

2.6.5 Robust yet Flexible

As the SLTT group discussed the philosophy and rules of science language for geologic-map databases, we came to understand that such language needs to accommodate tensions that exist between two competing requirements: (1) the need to be rigorous and robust, and (2) the need to allow the geologic-map maker to uniquely describe and interpret earth materials that occur within the map area. This tension boils down to the battle between robustness and flexibility:

- To be *robust*, language definitions must be clear and unambiguous, and parent-child relations among categories must be logical and based on common sense;
- To be *flexible*, the classification structure should not paint the data-producer into a corner at high levels of the classification: the schema must allow the field geologist and map compiler to move fairly deep into the classification hierarchies before feeling constrained by narrowly-defined terms whose meaning might be more stringent than the data-producer intends.
- To be *even more flexible*, the classification structure should allow the mapping geologist to use the SLTT science-language standard to build a "local favorites list" using concepts and terms defined in the standard. Thus, even though terms like "black shale" or "greenstone" or "mangerite" may not be defined in the standard, a local favorites list could contain these terms mapped into the SLTT science-language structure in the following fashion (Table 2.6.5.1):

Table 2.6.5.1 Rules and procedures for building "local favorites list"

Local term	Local meaning	SLTT concept	SLTT concept	SLTT concept
black shale	Fissile claystone containing abundant organic matter	*grain size* (specify clay:silt:sand ratio)	*depositional fabric* (specify fissile fabric)	*composition* (specify amount and type of organic content)
greenstone	Lower-greenschist facies mafic to intermediate rock	*Metamorphic facies*	fabric (specify fabric)	*composition* (specify composition)
mangerite	A charnockitic plutonic rock equivalent to an orthopyroxene-bearing monzonite	*modal mineralogy* (specify pyroxene modal percent)	*family* (specify monzonite)	*genesis* (specify plutonic igneous)

The only catch to this flexibility is that lithology terms in a "local favorites list" must be formally defined using science concepts and language laid out in the SLTT standard, and using data fields equivalent to those defined in the NADM data-model standard.

2.6.6 Compliance with North American Traditions

Earth-material names and parent-child relations among them must make sense according to common North American practice.

2.6.7 Rock Names versus *Modifiers*

Classification schema must distinguish clearly between *defined earth-material names* versus *modifiers that add information to each name*. Distinctions between names and modifiers should be incorporated into the architecture of the data fields and relational tables that support geologic-map databases, rather than into the rock names themselves.

2.7 Does genesis play a role in Earth-material classification?

In general, SLTT does not use genesis as a basis for classifying earth materials. Obviously, at the highest classification levels, earth-material names reflect their genesis (e.g., the distinction between *igneous*, *metamorphic*, and *sedimentary* materials fundamentally reflects genesis). Even deeper-level categories reflect a genetic factor (e.g., the distinction between *terrigenous-clastic* and *carbonate* sedimentary materials, or the distinction between *volcanic* and *plutonic* igneous materials). Obviously, the origin and geologic history of earth materials are important to many geologic-map users, and should be recorded in the geologic-map database in appropriate tables and data fields. However, genesis is so interpretive that its use in taxonomic classification at deeper levels should be avoided. Moreover, many map users are interested in the physical characteristics of earth materials, not their genesis. Hence, SLTT avoids the use of genetic factors in its classification schema.

3 SLTT RESULTS

Here we summarize the higher-level architecture of the SLTT science-language construct. These results and their detailed underpinnings will be released for peer review by the broad North American geologic-mapping community following final authorization by NADMSC.

3.1 Science language for earth materials

In parallel with the NADMSC Data Model Design Team (DMDT), SLTT defines the highest level in the classification hierarchy as "earth material":

Earth material—A naturally occurring substance formed in or on the Earth by physical, chemical, or biogenic processes that produce solid particles or crystals of mineral and (or) rock[3].

SLTT organizes earth materials into the following hierarchy:

- Earth Material
  - Igneous earth material
    - Hypabyssal rock
    - Plutonic rock
    - Volcanic rock
  - Composite-genesis earth material
    - Metamorphic rock (traditional usage)
    - Cataclastic rock
    - Impact-metamorphic material
  - Sedimentary earth material
    - Unconsolidated sediment
    - Consolidated sedimentary rock

3.2 Igneous materials

The science language of igneous materials was addressed by two SLTT subgroups, one dealing with *volcanic igneous materials* and the other dealing with *plutonic igneous materials*. In one sense this subdivision is arbitrary, as the processes, compositions, and textures of the two igneous families overlap. However, the accumulation of volcanic materials at the Earth's surface yields geologic products having unique geomorphic, compositional, and textural attributes; accordingly, SLTT developed science language for volcanic materials separately from plutonic materials.

3.2.1 Volcanic igneous materials

SLTT science language for volcanic earth materials is structured around four concepts:

---

[3] This is similar to the DMDT definition of *earth material*: "the substance of the solid Earth (rocks, minerals, organic material, glass, void space), defined based on intrinsic properties independent of their disposition within the Earth" (North American Data Model Steering Committee, 2003).

- Material name based on *genesis*
  - Intrusive
  - Effusive
  - Volcaniclastic
- Material name based on *texture*
  - Unconsolidated
  - Consolidated
    - Fragmental volcanic rock
    - Non-fragmental volcanic rock
- Material name based on *modal composition*
  - Felsic volcanic material
  - Mafic volcanic material
  - Ultramafic volcanic material
  - High-alkali volcanic material
  - Volcanic carbonatite material
  - Lamprophyre material
    - Fragmental volcanic rock
    - Non-fragmental volcanic rock
- Form
  - Intrusive
  - Constructional

Deeper-level classification categories apportion material names recommended by the IUGS (IUGS, 1973; Streckeisen, 1974, 1976, 1978, 1979; Le Bas and others, 1986; Le Maitre and others, 1989; Le Bas and Streckeisen, 1991).

3.2.2 Plutonic igneous materials

SLTT science language for plutonic earth materials generally adopts the British Geological Survey (BGS) classification scheme for plutonic rocks (Gillespie and Styles, 1999) in accordance with material names recommended by the IUGS (IUGS, 1973; Streckeisen, 1974, 1976, 1978, 1979; Le Bas and others, 1986; Le Maitre and others, 1989; Le Bas and Streckeisen, 1991). However, the SLTT scheme differs slightly in order to accommodate North American traditions.

3.3 Composite-genesis rocks and rock particles

As defined in the SLTT classification, composite-genesis earth material is any earth material having observable features that document mineralogical, chemical, or structural change of a preexisting earth material essentially in the solid state. The category includes metamorphic rocks (*sensu strictu*), hydrothermally altered rocks, cataclastic rocks, and impact-metamorphic rocks. Weathered rock and pedogenic soil also could be considered composite-genesis materials, but SLTT has not included these materials in the development of the classification. Where possible, the British Geological Survey classification of metamorphic rocks (Robertson, 1999) and preliminary recommendations of the IUGS Subcommission on the Systematics of Metamorphic Rocks (SCMR) (Schmid and others, 2002) were adapted to meet SLTT database requirements.

Composite-genesis rocks are classified along two orthogonal dimensions, fabric and composition. Because both of these dimensions are hierarchical (Figures 1 and 2), the class hierarchy for composite-genesis rocks is a directed acyclic graph rather than a tree. Classes that have

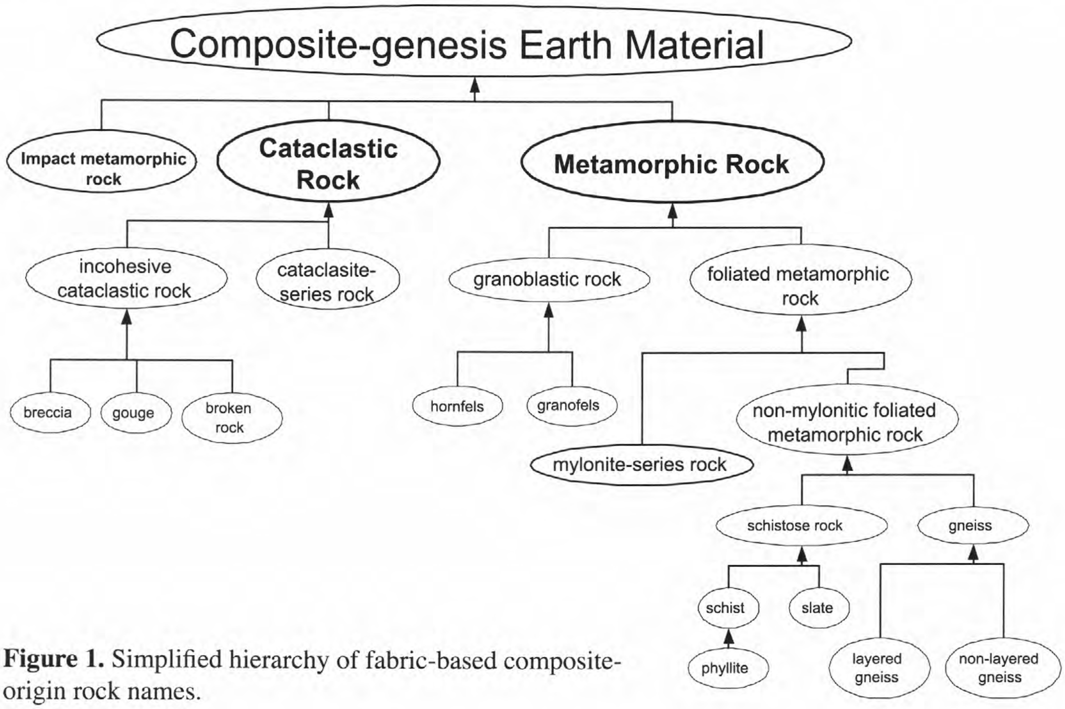


Figure 1. Simplified hierarchy of fabric-based composite-origin rock names.

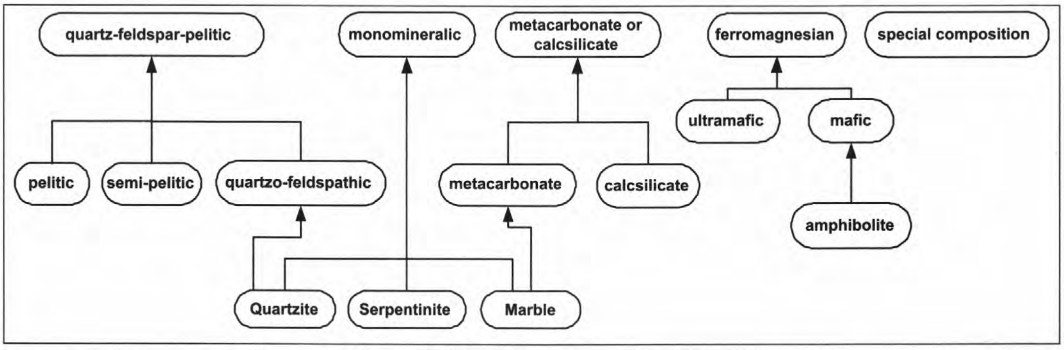


Figure 2. Simplified hierarchy of composition-based composite-origin rock name terms.

both composition and fabric criteria are 'children' of both a 'composition' parent and a 'fabric' lithology parent. Actual class names for rocks have a *fabric* component (such as schist) and a *compositional* component (such as marble or quartzofeldspathic).

Metamorphic rock (traditional sense)—A metamorphic rock has observable features that document change after the original formation of the rock, under physical or chemical conditions that differ from those normally occurring at the surface of the Earth and in zones of cementation and diagenesis below the surface (Smulikowski and others, 1997). The basic level of classification is the definition of fabric-based types described as hornfels, granofels, schist, layered gneiss, non-layered gneiss, slate, and mylonite-series rock (Table 3.3.1). In this classification, hydrothermally altered rocks are treated as metamorphic rocks; the full metamorphic classification is applied to these rocks without treating them as a distinct, separate category.

Table 3.3.1 Fabric-based metamorphic-rock classes

Fabric-based Metamorphic-rock Class	**Definition and scope notes**
Hornfels	A non-foliated aphanitic metamorphic rock having granoblastic fabric. The term does not necessarily denote a contact metamorphic origin, although that is most commonly the case
Granofels	A phaneritic metamorphic rock that has little or no foliation or lineation, implying that that less than 10% of the particles in the rock are fabric elements that have an inequant shape and an aspect ratio ≥ 1.5:1
Schist	a phaneritic metamorphic rock that has well developed continuous schistosity. "Well developed" schistosity is defined here to mean that >50% of the rock consists of mineral grains having a tabular, lamellar, or prismatic crystallographic habit that are oriented in a continuous planar or linear fabric (Jackson, 1997). Continuous is defined to mean that domains lacking the fabric are <1 cm thick if they are layers, and <5 cm in diameter if they are irregular patches, and constitute <25% of the rock. Phyllite is a fine-grained subclass of schist having an average grain size between 0.25 mm and 0.1 mm and a silvery sheen.
Layered gneiss	a foliated, phaneritic rock that lacks well developed, continuous schistosity and has laterally continuous compositional layering > 5 mm thick
Non-layered gneiss	a foliated, phaneritic rock that lacks both well developed, continuous schistosity and laterally continuous compositional layering > 5 mm thick. "Laterally continuous" here means that layers defining the foliation can be traced > 10 cm (length of lateral continuity)
Slate	an aphanitic rock that has well developed schistosity (Brodie and others, 2002). The definition of schistosity used in this classification requires that >50% of the rock consists of mineral grains having a tabular, lamellar, or prismatic crystallographic habit that are oriented in a continuous planar or linear fabric. In an aphanitic rock this determination is generally based on indirect evidence, which is typically the presence of slaty cleavage, and the sheen observed on parting surfaces due to alignment of tiny phyllosilicate grains. An average grain size that is aphanitic (<0.1 mm), except for porphyroblasts, is specified for precision.
Mylonite-series rock	displays a foliation defined by the shapes of deformed mineral grains or grain aggregates having aspect ratios > 1.5:1, >10% of the rock is composed of "matrix" showing evidence of tectonic reduction in grain size, and the foliation and matrix have observable features that document continuous, crystal-plastic deformation. Mylonite-series rock is subdivided according to matrix percentages into protomylonite, mylonite, and ultramylonite (Sibson, 1977)

Composition-based rock classes include amphibolite, marble, and common monomineralic rocks such as quartzite and serpentinite, which are defined individually by modal mineral composition. Other monomineralic metamorphic rocks consisting predominantly (>75%) of a single mineral are classified as monomineralic-granofels, monomineralic-hornfels, monomineralic-schist, etc., depending on the fabric. Composition qualifiers defined on the basis of modal mineralogy (ferromagnesian, calcsilicate, carbonate, pelitic, semipelitic, quartzofeldspathic; see Table 3.3.2) are combined with a fabric term as in "pelitic schist." Traditional non-systematic rock terms such as amphibolite that are based on modal mineralogy are also treated as composition qualifier terms. This classification leans towards a minimum of special rock names for unusual composition rocks, and such rocks would be assigned a 'special composition' qualifier. The uncontrolled rock-name field in the database is available to assign any special rock name the geologist may prefer.

Table 3.3.2 Selected composition-based classes (qualifiers)

Composition qualifiers	Definition
amphibolite	Rock consists of >75% green, brown, or black amphibole plus plagioclase (including albite) and amphibole >30% (modal) of whole rock, and amphibole >50% of total mafic constituents
argillic	Rock is apparently clay-rich. Use for aphanitic rocks
calcareous	Rock reacts to form bubbles when hydrochloric acid is applied. Use for aphanitic rocks (e.g. hornfels)
calcsilicate	Rock consists of ≥ 50% calcsilicate or carbonate minerals and carbonate minerals ≤ calcsilicate minerals in mineral mode
ferromagnesian	Rock consists of >40% dark ferromagnesian minerals. Standard term defined by Bates and Jackson (1987) to mean "containing iron and magnesium"
impure marble	Rock consists of >50% calcsilicate or carbonate minerals and relative proportion of calcsilicate and carbonate minerals is unknown or not specified
mafic	Rock consists of ≥ 40% and <90% ferromagnesian minerals
marble	Rock consists of > 75% carbonate minerals
metacarbonate	Rock consists of >50% calcsilicate or carbonate minerals and carbonate minerals > calcsilicate minerals in mineral mode
monomineralic	Rock consists of >75% of a single mineral species and does not meet any of the other composition terms (e.g. quartzite, calcite marble, dolomite marble, serpentinite…)
pelitic	Rock for which the sum of modal quartz+feldspar+ mica + aluminous mineral is ≥ 70%, and aluminous mineral + mica content is ≥ 40%
quartzite	Rock consists of ≥ 75% quartz
quartzofeldspathic	Rock for which the sum of modal quartz+feldspar+ mica + aluminous mineral is ≥70%, and quartz + feldspar (sensu Robertson, 1999) >60%
semipelitic	Rock for which the sum of modal quartz+feldspar+ mica + aluminous mineral is ≥70%, and quartz+ feldspar < 60%
silicic	Rock is apparently silica-rich. Use for aphanitic rocks. Bates & Jackson (1987) include denotation of igneous origin. For this classification, should be considered to mean "appears to consist largely of quartz and feldspar", generally is aphanitic with hardness >= 6.
special composition	Rock has a mineral composition that doesn't fit in any defined composition class. A modal mineral description is essential. The rock consists of <40% ferromagnesian minerals and <50% carbonate + calcsilicate minerals and <70% Q+Fs + mica + aluminous minerals
ultramafic	Rock consists of >90% ferromagnesian minerals

The SLTT classification of metamorphic rocks does not apply the 'meta' prefix to a protolith name (as in metasiltstone), because it cannot be simply integrated into a classification based on fabric and composition, and because interpretations of protolith can be highly subjective. Rock name terms in the form 'meta-(*some rock name*)' can be placed by the user in an uncontrolled rock name field and can also appear in a user interface by having underlying software that maps the name assignment to the implied dual classification. Where the protolith can be determined, the classification includes two distinct parts—a protolith classification using the criteria applicable to the protolith lithology, and a composite-genesis classification based on the fabric and composition criteria outlined here. The data model design should include a mechanism that allows the 'dominant' aspect of a rock that has multiple classifications to be specified.

Cataclastic rock—SLTT classifies a rock as cataclastic if >10% of the volume consists of fragments bounded by fractures. Cataclastic rocks are further classified based on the presence or absence of primary cohesion, the percentage of broken fragments large enough to be visible, and the amount of fragmental cataclastic matrix (Sibson, 1977; Snoke and Tullis, 1998) Cataclastic rock having evidence of primary cohesion is subdivided according to matrix percentages into *protocataclasite*, *cataclasite*, and *ultracataclasite* (Sibson, 1977). Cataclastic rock that lacks primary cohesion is subdivided into fault breccia (visible fragments >30% of rock) and gouge (visible fragments <30%).

Impact-metamorphic rock—Impact-metamorphic rocks have observable features, such as microscopic planar deformation features, that are unequivocally the result of shock metamorphism (Stöffler and Grieve, 2001), high-pressure minerals, or field evidence such as shatter cones and crater structure. Adapting Stöffler and Grieve's (2001) IUGS recommendations with slight modifications, impact-metamorphic rock is classified as shocked rock, impact melt rock, or impact breccia.

3.4 Sedimentary materials

The SLTT Sedimentary Subgroup developed science language for the lithologic classification (material name), physical characteristics, and origin and depositional history of sedimentary materials.

3.4.1 Classification

At the top hierarchical level, sedimentary earth materials are classified into eight categories based on sediment composition (Table 3.4.1.1). At a high level, unconsolidated sediment is distinguished from consolidated rock, and separate (but parallel) naming schemes are developed for each consolidation state.

Table 3.4.1.1 Higher taxonomic levels of sedimentary-material classification

Sedimentary Material (unconsolidated, consolidated)	Definition
Sedimentary material, unclassified	Not enough information is known about a sedimentary material to classify it as anything other than sedimentary rock or sediment
Terrigenous-clastic sedimentary material	A rock or sediment composed principally of broken fragments that are derived from the land or continent. To be considered as terrigenous-clastic, a rock (sediment) must have ≥ 50% of its constituents derived from the land or continent
Carbonate sedimentary material	Sediment or sedimentary rock ≥50% of whose primary and (or) re-crystallized constituents are composed of carbonate minerals (calcite, aragonite, dolomite). By definition, such materials are *intra-basinal* in origin—that is, they formed by processes operating within the depositional regime, and were not transported into that regime from other sediment sources
Organic-rich sedimentary material	Sedimentary materials having sufficiently high organic content that they can not be identified as another kind of sedimentary rock (e.g., terrigenous-clastic or carbonate). Pragmatically, SLTTS_1.0 places this threshold at ≥50% organic content by weight to be consistent with the established definition for coal without conflicting with definitions of other compositionally-based categories
Non-clastic siliceous sedimentary material	Sedimentary materials dominated by non-clastic silica are those composed of ≥50% silica of biogenic or chemical origin (Hallsworth and Knox, 1999, p. 21)
Noncarbonate-salt sedimentary material	Sedimentary materials dominated by non-carbonate salts are those whose primary constituents consist of chloride, sulphate, or borate minerals. Such materials also are known as *evaporite* materials because they form through evaporative precipitation of mineral salts from brines—either directly from the water column or from pore fluids during diagenesis
Phosphate-rich sedimentary material	Phosphatic sedimentary materials are those in which phosphate minerals or phosphatic components comprise >50% of the sedimentary framework as determined by hand-lens or petrographic analysis (Hallsworth and Knox, 1999). This corresponds with a rock (sediment) typically containing ≥ 15% P_2O_5 (by weight)
Iron-rich sedimentary material	Iron-rich sedimentary materials are those in which iron-bearing minerals comprise ≥ 50% of the sedimetary framework as determined by hand-lens or petrographic analysis (Hallsworth and Knox, 1999). This corresponds with a rock (sediment) typically containing ≥ 15% iron (by weight)

Within each compositional category, lower-level (more detailed) material names are based on textural or compositional criteria (or both), depending on the parent category.

3.4.2 *Unconsolidated versus consolidated sedimentary materials*

The distinction between "unconsolidated" and "consolidated" (sediment *versus* rock) occupied considerable SLTT discussion and attention. We concluded that SLTT can suggest guidelines for distinguishing unconsolidated from consolidated materials. However, ultimately it will be the subjective decision of the data producer as to whether a specific sedimentary material is "consolidated" or "unconsolidated" according to his or her judgment. Table 3.4.2.1 provides guidelines that can facilitate this determination.

3.4.3 *Outcrop characteristics of sedimentary materials*

SLTT developed science language for a variety of attributes that characterize the outcrop appearance of sedimentary materials (Table 3.4.3.1):

Table 3.4.2.1 Degrees of consolidation (modified from Bowles, 1984, Table 5-2)

	Lithification State	**Field Criterion**	**Relative Density (D_r)**[4]
Unconsolidated	Very slightly consolidated	Easily indented with fingers	0.00—0.20
	Slightly consolidated	Somewhat less easily indented with fingers. Easily shoveled	0.20—0.40
	Moderately consolidated	Shoveled with difficulty	0.40—0.70
Consolidated	Well consolidated	Requires pick to loosen for shoveling	0.70—0.90
	Lithified	Requires blasting or heavy equipment to loosen	0.90—1.00
	Indurated	Rings to the blow of a hammer	1.00

[4] As translated by Bowles (1984, p. 151-152), relative density is an engineering parameter that relates void space determined in the laboratory to a ratio involving index values of minimum and maximum void space for specified materials under specified conditions. Void space in turn is related to *in situ* dry unit weight. Also see the *Glossary of Geology* definition of relative density (Jackson, 1997, p. 540).

Table 3.4.3.1 Outcrop characteristics of sedimentary materials

Science Concept	**Data-Field Content**
Lithotope abundance	Indicates the relative abundance of each lithotope in an outcrop
Map-unit geomorphology	Describes how a map unit crops out (prominent, subdued)
Outcrop profile	Describes how individual lithotopes crop out (ledge-forming, slope-forming, etc.)
Outcrop weathering	Describes how sedimentary materials weather (cavernous, friable, etc.)
Upper-surface geomorphology	Describes various features associated with the upper surface of unconsolidated materials (dissection, pavement, pedogenic soils, clast weathering, etc.)
Material color (fresh)	Describes the color of fresh geologic materials
Material color (weathered)	Describes the color of weathered geologic materials
Material color (dry)	Describes the color of dry geologic materials
Material color (wet)	Describes the color of wet geologic materials
Consolidation state	Describes how firm and knitted together a sedimentary material is, and how hard it is once it has been lithified

3.4.4 Sedimentary structures

The SLTT classifies sedimentary structures into primary and secondary structures:

Primary sedimentary structure—A sedimentary structure, either inorganic or organic, formed during the accumulation and (or) penecontemporaneous modification of sediment, *intrinsically reflecting conditions of the sedimentary environment.*

A primary structure reflects one of the following processes:

(1) transportation and deposition of sediment (i.e., mechanical process);
(2) penecontemporaneous modification of sediment by physical and biogenic processes shortly after its accumulation (e.g., downslope slumping, bioturbation, burrowing);

Secondary sedimentary structure—A sedimentary structure "...that originated subsequently to the deposition...of the rock in which it is found...; esp. an epigenetic sedimentary structure, such as a concretion or nodule produced by chemical action, or a sedimentary dike formed by infilling" *Glossary of Geology* (Jackson, 1997, p. 576).

Not included as secondary sedimentary structures are:

(1) structures representing from tectonic deformation (e.g., faults, fractures)
(2) structures formed as the result of subaerial exposure, weathering, and dissolution not related to the sedimentary life cycle of a map unit (e.g., karsts, pedogenic soils).

The essential ingredient of secondary sedimentary structures is that they *have no intrinsic relation with conditions of the sedimentary environment.* This distinguishes secondary structures from primary structures.

Table 3.4.4.1 summarizes major categories of sedimentary structures.

Table 3.4.4.1 Classification of sedimentary structures

Primary Structures
- Inorganic structures
  - Syngenetic structures
    - Depositional structures
    - Erosional structures
  - Penecontemporaneous structures
    - Bed-surface structures
    - Within-bed structures
    - Multi-bed structures
- Biogenic Structures

Secondary Structures
- Secondary deformation structures
  - Sedimentary hardground
  - Dissolution structures
  - Epigenetic growth structures

Uclassified Structures
- Bed-surface structures
- Within-bed structures
- Multi-bed structures

3.4.5 Fabric and texture of sedimentary materials

SLTT developed science language for a variety of attributes that characterize the fabric and texture of sedimentary materials (Table 3.4.5.1).

Table 3.4.5.1 Fabric and texture elements for sedimentary materials

Science Concept	Data-Field Content
Preservation of depositional fabric	Yes or no
Grain-support *versus* matrix support	Indicates whether fabric is clast-supported or matrix-supported
Particle size, matrix grain size (range)	Indicates range of matrix grain size
Particle size, matrix grain size (average)	Indicates mean of matrix grain size
Particle size, particle grain size (range)	Indicates range of particle grain size
Particle size, particle grain size (average)	Indicates mean of particle grain size
Particle sorting	Indicates sorting in terms of Inclusive Graphic Standard Deviation (Folk, 1968)
Particle shape and rounding	Indicates the shape of grains and clasts (rounded, subangular, tabular, spherical, etc.)
Coated particles	Indicates the fabric type created by particle coating (ooidal, pisoidal, oncoidal)
Particle orientation	Indicates the geometric orientation of elongate or disk-shaped particles
Particle packing	Indicates the spacing or density patterns of particles as expressed by nature of grain contacts

3.4.6 Sedimentary genesis

SLTT classifies sedimentary genesis according to a scheme that integrates three attributes:

- *Depositional process* (the mechanism by which a sediment is transported and deposited);
- *Depositional environment* (the conditions under which sediment is deposited, as defined by ambient physical, chemical, and biological conditions);
- *Depositional product* (the three-dimensional physiographic and geomorphic deposit type resulting from a process operating within an environment).

Ultimately, it is the interaction between *depositional process* and *depositional environment* that yields a *depositional product*. This interaction can be viewed as a two-dimensional matrix in which process is arrayed against place (Table 3.4.6.1).

Table 3.4.6.1 Two-dimensional matrix that arrays *depositional process* (left) against *depositional environment* (right) to yield cells in which a *depositional product* may or may not occur (for examples, see Table 3.1.4). NOTE: Table 3.4.6.1 depicts only the highest-level categories for process and environment; the matrix can be expanded to expose deeper and deeper categories. Also Note: Not all combinations of process and environment are possible (hence, no depositional product exists).

					ENVIRONMENTS (major environments only)		
					Terrestrial	**Marginal marine**	**Marine**
PROCESSES (major processes only)	**Inorganic Processes**	Chemogenic Processes	Evaporative precipitation				
			Seafloor precipitation				
		Fluid-Flow	Aqueous-flow	current flow			
				fluvial flow			
				overland-flow			
				wave action			
			Eolian-flow	dune- and sand-sheet processes			
		Gravitational Potential Energy	Mass-wasting	subaerial			
				subaqueous			
			Particle Settling	object plummeting			
				suspension deposition			
				Impact-ejecta accumulation			
		Solid-flow	Ice-flow				
	Biogenic Processes	Sediment-production	biogenic oozes				
		Framework-building	framework-built reefs				
		Sediment-binding	organic mats				
		Sediment-trapping	organic-baffles and reef mounds				

In Table 3.4.6.1, the intersection of a *genetic process* with a *sedimentary environment* yields a grid cell that represents a potential *depositional product.* Table 3.4.6.2 lists representative examples (but not all examples) of such products.

Table 3.4.6.2 Representative examples of deposit types that represent the intersection of depositional process and depositional environment (see Table 3.4.6.1)

algal-mat deposit	bog deposit	bar deposit	beach deposit	braided-channel deposit	channel deposit	chute-channel deposit	crevasse-channel deposit
crevasse-splay deposit	debris-flow deposit	distributary-mouth bar deposit	dune deposit	fan deposit (subaerial)	fan deposit (subaqueous)	fan-delta deposit	flood-plain deposit
glacial-till deposit	inlet channel fan deposit	lagoon deposit	levee deposit (subaerial)	levee deposit (subaqueous)	marsh deposit	meandering-channel deposit	mud-flat deposit
overbank-fines deposit	pelagic-ooze deposit	pond deposit	reef, framework-built	reef, sediment-trapping	sabkha deposit	sand-flat deposit	sand-flat deposit
sheet-flow deposit	sheet-sand deposit	shelf deposit	slide deposit	slope deposit	slump deposit	supratidal-flat deposit	swamp deposit
tidal-channel deposit	tidal-flat deposit	tidal-inlet deposit	tidal-ridge deposit	turbidite deposit	washover-fan deposit		

All three genetic attributes (depositional process, depositional environment, depositional product) can be classified hierarchically to yield a complete description of how a sedimentary material was formed.

4 PEDOGENIC MATERIAL

SLTT has not yet found a satisfactory basis for classifying pedogenic-soil materials. A pedogenic soil is a parent material that has developed soil structure, texture, and mineralogy (horizonation) through a variety of physical, chemical, and biologic agents. Horizon development in the parent materials thus is a secondary process, not a primary process that leads to a first-generation earth material. Even though pedogenic processes take place at temperatures, pressures, and physical-chemical environments at the Earth's surface, soil-forming processes are akin to "metamorphism" in the sense that an original earth material (parent material) is transformed into another earth material (pedogenic soil). Thus, pedogenic material could be included within the composite-genesis category. However, some SLTT members argue that pedogenic soil properly belongs within the sedimentary category, while others maintain that pedogenic soil is a high-level taxonomic category comparable to igneous, sedimentary, and composite-genesis materials. SLTT hopes to resolve these conflicting interpretations by the time the science-language documents are released for widespread North American peer review.

5 CONCLUSIONS

The use of standardized science language in digital geologic-map databases is a new frontier that is likely to evolve with time and experience. With this in mind, we are developing a classification of earth materials that we believe reflects not only how mapping geologists view them but also how such materials might be queried and analyzed in digital geologic-map databases. No single classification of earth materials is going to please all workers. However, the schemes we propose hopefully will be clearly understandable, internally consistent, and usable by both data-producer and data-user.

6 REFERENCES CITED

Bowles, J.E., 1984, Physical and geotechnical properties of soils: New York, McGraw-Hill Book Company, 2nd Edition, 578 p.

Brodie, K., Fettes, D., Harte, B., and Schmid, R., 2002, Towards a unified nomenclature of metamorphic petrology: 5. Structural terms including fault rock terms, A proposal on behalf of the IUGS Subcommission on the Systematics of Metamorphic Rocks, Web version of 31.10.2002: 8 p., available online at http://www.bgs.ac.uk/SCMR

Dreimanis, A., 1988, Tills: their genetic terminology and classification, *in* Goldthwait, R.P., and Matsch, C.L., eds., Genetic classification of glacigenic deposits (Final report of the Commission on Genesis and Lithology of Glacial Quaternary Deposits of the International Union for Quaternary Research [INQUA]): Rotterdam, A.A. Balkema, p. 17-83.

Foley, S.F., Venturelli, G., Green, D.H., and Toscanni, L., 1987, The ultra potassic rocks: characteristics, classification and constraints for petrographic models: Earth Science Reviews, v. 24, p. 81–134.

Folk, R.L., 1968, Petrology of sedimentary rocks: Austin, Texas, Hemphill's Bookstore, 170 p.

Gillespie, M.R., and Styles, M.T., 1999, Classification of igneous rocks: Volume 1 of the BGS rock classification scheme: British Geological Survey Research Report Number RR 99-06, 52 p.

Goldthwait, R.P., and Matsch, C.L., eds., 1988, Genetic classification of glacigenic deposits (Final report of the Commission on Genesis and Lithology of Glacial Quaternary Deposits of the International Union for Quaternary Research [INQUA]): Rotterdam, A.A. Balkema, 294 p.

Goldthwait, R.P., Matsch, and Dreimanis, A., 1988, Preface, *in* Goldthwait, R.P., and Matsch, C.L., eds., Genetic classification of glacigenic deposits (Final report of the Commission on Genesis and Lithology of Glacial Quaternary Deposits of the International Union for Quaternary Research [INQUA]): Rotterdam, A.A. Balkema, p. vii-ix.

Hallsworth, C.R., and Knox, R.W.O'B., 1999, Classification of sediments and sedimentary rocks: Volume 3 of the BGS rock classification scheme: British Geological Survey Research Report Number RR 99-03, 44 p.

Heiken, G., and Wohletz, K.H., 1985, Volcanic ash: Berkeley, University of California Press.

International Union of Geological Sciences Subcommission on the Systematics of Igneous Rocks, 1973, Plutonic rocks: Geotimes, v. 18, no. 10, p. 26-30.

Jackson, J.A., 1997, Glossary of geology, 4th ed.: Alexandria, Virginia, American Geological Institute, 769 p.

Le Bas, M.J., and Streckeisen, A., 1991, The IUGS systematics of igneous rocks: Journal of the Geological Society of London, v. 148, p. 825–833.

Le Bas, M.J., Le Maitre, R.W., Streckeisen, A., and Zanettin, B., 1986, A chemical classification of volcanic rocks based on the total alkali-silica diagram: Journal of Petrology, v. 27, p. 745–750.

Le Maitre, R.W., and 11 others (editors), 1989, A classification of igneous rocks and glossary of terms: Recommendations of the International Union of Geological Sciences Subcommission on the Systematics of Igneous Rocks: Oxford, Blackwell Scientific Publications.

Le Maitre, R.W. (editor), 2002, Igneous rocks, A classification and glossary of terms, second edition: Cambridge, Cambridge University Press, 2nd edition 236 p.

MacDonald, R., 1974, Nomenclature and petrochemistry of the peralkaline oversaturated extrusive rocks: Bulletin Volcanologique, v. 38, p. 498–516.

McMillan, A.A., and Powell, J.H., 1999, Classification of artificial (man-made) ground and natural superficial deposits: applications to geological maps and datasets in the UK: v. 4 of the BGS rock classification scheme: British Geological Survey Research Report Number RR 99-04, 65 p.

North American Commission on Stratigraphic Nomenclature, 1983, North American stratigraphic code: American Association of Petroleum Geologists Bulletin, v. 67, no. 5, p. 841-875.

North American Data Model Steering Committee, 2003, Proposed North American standard conceptual model for geologic map information, version 1.0: U.S. Geological Survey Open-File Report 04-XXX (http://)

Robertson, S., 1999, Classification of metamorphic rocks: Volume 2 of the BGS rock classification scheme: British Geological Survey Research Report Number RR 99-02, 24 p.

Schmid, R., 1981, Descriptive nomenclature and classification of pyroclastic deposits and fragments: Recommendations of the IUGS Subcommission on the Systematics of Igneous Rocks: Geology, v. 9, p. 41–43.

Schmid, R., Fettes, D., Harte, B., Davis, E., Desmons, J., and Siivola, J., 2002, Towards a unified nomenclature in metamorphic petrology: 1. How to name a metamorphic rock, A proposal on behalf of the IUGS Subcommission on the Systematics of Metamorphic Rocks, Web version of 31.07.2002: 13 p., available online at http: //www.bgs.ac.uk/SCMR

Sibson, R.H., 1977, Fault rocks and fault mechanisms: Journal of the Geological Society of London, v. 133, p. 191-213.

Smulikowski, W., Desmons, J., Harte, B., Sassi, F.P., and Schmid, R., 1997, Towards a unified nomenclature of metamorphism: 3. Types, grade and facies [A proposal on behalf of the IUGS Subcommission on the Systematics of Metamorphic Rocks], compiled 1 September 1997 and corrected 19 November 1997: 13 p., available online at http: //www.bgs.ac.uk/SCMR.

Snoke, A.W., and Tullis, J., 1998, An overview of fault rocks, *in* Snoke, A.W., Tullis, J., and Todd, V.R., Fault-related Rocks, A Photographic Atlas: Princeton, N.J., Princeton University Press, p. 3-18.

Stöffler, D., and Grieve, R.A.F., 2001, IUGS classification and nomenclature of impact metamorphic rocks: Towards a final proposal [indexed as: Towards a unified nomenclature of metamorphism: IUGS classification and nomenclature of impact metamorphic rocks. A proposal on behalf of the IUGS Subcommission on the Systematics of Metamorphic Rocks, Web version of 14.01.2001]: 3 p., available online at http://www.bgs.ac.uk/SCMR.

Streckeisen, A. 1974, Classification and nomenclature of plutonic rocks: Recommendations of the IUGS Subcommission on the Systematics of Igneous Rocks: Geologische Rundschau Internationale Zeitschrift für Geologie: Stuttgart, v. 63, p. 773–785.

Streckeisen, A. 1976, To each plutonic rock its proper name: Earth Science Reviews, v. 12, p. 1–33.

Streckeisen, A. 1978, IUGS Subcommission on the Systematics of Igneous Rocks: Classification and nomenclature of volcanic rocks, lamprophyres, carbonatites and melilitic rocks. Recommendations and suggestions: Neues Jahrbuch für Mineralogie, v. 134, p. 1–14.

Streckeisen, A. 1979, Classification and nomenclature of volcanic rocks, lamprophyres, carbonatites and melilitic rocks: recommendations and suggestions of the IUGS Subcommission on the Systematics of Igneous Rocks: Geology, v. 7, p. 331–335.

Webster's Encyclopedic Unabridged Dictionary of the English Language, 2001: San Diego, Thunder Bay Press, 2230 p.

APPENDIX 1

PROPOSED NORTH AMERICAN GEOLOGIC-MAP DATA MODEL SCIENCE-LANGUAGE TECHNICAL TEAM CHARTER

11/1/99

Executive Statement.—See the Data Model Steering Committee (DMSC) charter for an executive statement on technical teams.

MANDATE AND CHARGE

Mandate.—The Science Language Technical Team (SLTT) is mandated to develop standardized nomenclature for digital geologic-map databases—including (but not limited to) the following areas:

- nomenclature for the description of geologic map units (lithology, stratigraphy, geomorphology, pedology, petrology, genesis, etc.)
- nomenclature for the description of linear geologic features (contacts, faults, fold axial traces, mapped marker units, geomorphic features, etc.),
- nomenclature for the description of point geologic features (structural points, etc.);
- nomenclature for descriptive and interpretive information about spatial and geologic relations among geologic map units, linear features, and point features (e.g., sequencing relations, stratigraphic relations, and geometric relations, etc.).

The standardized terminology will support a proposed standard geologic-map data model for North America.
Charge.—To achieve its mandate, the Science Language Technical Team is charged with the following tasks:

(1) To determine the scope and comprehensiveness appropriate to a continent-wide terminology for geologic map databases. Terminology scope should reflect several realities, including (1) the intended use of the geologic-map terminology, (2) the geologic scale to which the terminology will be applied, (3) the prerogatives of historic usage by various geologic-mapping constituencies, and (4) the degree to which geologic terminology is amenable to a single hierarchical classification structure. These factors (and others developed by the SLTT) should determine the degree and level of standardization appropriate for continent-wide geoscience language.
(2) To develop one or more strawman classifications for geologic-map science language that will be made available for widespread peer review.
(3) To prepare and publish documents describing the basis for the science-language terminology, and presenting the classification scheme(s) and their technical and non-technical definitions.

Authority.—The SLTT derives its authority and legitimacy from the DMSC, which provides guidance and requirements on behalf of the constituencies it represents.

Accountability.—The SLTT is accountable to the DMSC. Through a representative mutually acceptable to the SLTT and DMSC, the SLTT periodically apprises the Steering Committee of progress toward science language terminology and about issues and problems that need consideration by the DMSC.

TECHNICAL-TEAM OPERATIONS

Execution of work.—The SLTT will convene an initial meeting to evaluate goals and to discuss issues, problems, and terminology strategies. The Technical Team should have as many face-to-face meetings as required to allocate responsibilities and to resolve issues and problems not easily resolvable via e-mail.

Lateral Coordination.—The SLTT will regularly communicate strategies and proposed terminologies laterally to other Technical Teams—especially the Data-Model Design Technical Team—in order to ensure that data-model architecture and software tools consistently reflect the evolving science language and concepts.

Technical Review.—Science-language documents prepared by the SLTT will be presented to the DMSC for initial review and evaluation for compliance with the overall goals of the North American geologic-map data model. Following DMSC review and SLTT response, the science-language documents will be widely distributed for technical peer review by the geosciences community (probably through a web-based venue).

TECHNICAL-TEAM MEMBERSHIP

Work Group Size.—The size of the SLT should be commensurate with its mandate: If geologic and political realities require that the scope and content of data-model science language be generalized and narrow, then the size of the Technical Team should be small; however, if the scope and content of data model science language is to be comprehensive and detailed, then the size of the Technical Team should be large enough to ensure scientific comprehensiveness and consensus of the larger geologic community.

Scientific Breadth.—SLTT membership should span the range of surficial and bedrock geologic disciplines, including expertise in sedimentary, igneous, metamorphic, structural, stratigraphic, and geomorphic/pedogenic arenas. Experts on specific scientific disciplines can be added for short durations to address specific geologic issues that arise during SLTT deliberations.

Geographic Breadth.—SLTT membership should include a broad range of geographic representation so as to reflect provincial geologic usages.

Constituency Breadth.—SLTT membership should represent the constituencies that will contribute to geologic map databases—initially including the U.S. Geological Survey (USGS), the Association of American State Geologists (AASG), the Geological Survey of Canada (GSC), and the Canadian Provincial Surveys. Inclusion of industry and academic participants will depend on the narrowness or breadth of the science-language standards.

Appointment procedure.—SLTT members shall be appointed by the DMSC based on the recommendations of each constituency and considering the criteria defined in Scientific, Geographic, and Constituency Breadth:

- AASG recommendations will come from the AASG Digital Geologic Mapping Committee;
- GSC recommendations will come from that agency as appropriate to its internal selection procedures;
- Recommendations from the Canadian Provincial Surveys will come from those agencies consistent with their interest and appropriate to their internal selection procedures;
- Recommendations from the USGS will come from that agency as appropriate to its internal selection procedures.

Lifespan of Technical Team.—Continued existence of the SLTT as a standing committee responding to data-model science–language needs shall be at the discretion of the DMSC. The SLTT will remain intact during the review period, and shall respond to Steering Committee review and to peer review until such time as version 1.0 of the science-language classification is adopted for use in the draft standard data model.

MILESTONES

Within a year of convening its first session, the SLTT shall carry out its charge to produce one or more science-language strawman classifications. The SLTT receives guidance on milestones from the DMSC, evaluates their feasibility, and reaches targets in conjunction with DMSC.

APPENDIX 2

Memoranda from SLTT Chair (J.C. Matti) to SLTT committee members

Memorandum from SLTT Chair to SLTT committee members (4/03/2000)

Participants on the Science Language Technical Team For a Proposed North American Geologic-Map Data Model

04/03/2000

Science Language Technical Team colleagues:

By now, each of you hopefully is aware that you have been selected as a participant on a technical team (Science Language Technical Team, SLTT) tasked with coming to groups with how (or whether) a common set of geoscience terms can be developed for digital geologic-map data bases produced in North America. The SLTT is one of several parallel teams commissioned on behalf of a proposed North American Geologic-Map Data Model.

Background

A standardized data-base model for the input, storage, manipulation, retrieval, and analysis of digital geologic-map information is being developed by a consortium of interests, including the Association of American State Geologists (AASG), the U.S. Geological Survey (USGS), the Geological Survey of Canada (GSC), and the Canadian Provincial Surveys. The data model currently being evaluated was developed as a cooperative venture by the USGS, the GSC, and the AASG.

This model attained visibility through a series of workshops and through presentations at national GSA meetings. It developed as a likely candidate for a North American data-model standard, and over a period of time was revised and refined under the aegis of the AASG, the USGS, and the GSC. The data model can be found on the World Wide Web at http://geology.usgs.gov/dm/model/Model43a.pdf (version 4.3, Johnson and others, 1999). Continued development of a data-model standard is proceeding under the auspices of a multi-constituency North American Data Model Steering Committee (NADMSC, http://geology.usgs.gov/dm/steering/), which has commissioned the technical teams that are developing various aspects of the data model.

How did you come to be a participant in this process?

Scientists from the American state geological surveys were identified by the Association of American State Geologists. Participants from the USGS were nominated through a process that was coordinated through the three Regional Geologists and through the Geologic Division Program Coordinators. The two Canadian participants represent the Provincial surveys and the Geological Survey of Canada; hopefully, additional Canadian participants will be identified in the near future. I am in the process of recruiting a few representatives from the geologic-map using community within the U.S. Departments of Interior and Agriculture, one of whom has been identified. All of you are viewed as ideal for responding to the task before us.

Where is the SLTT process now, and what is next?

It has taken a while to establish the SLTT membership because it has not been easy to coordinate among multiple constituencies. But, we are about there, so I thought I would bring you up to speed, and address some mechanical issues.

(1) I want to start business on April the 17th.
(2) We have a year from that date to execute our responsibilities.
(3) I imagine a month of your time will be required throughout the 12-month cycle, but time invested will depend on interest level and commitment to the SLTT process.
(4) As a Team, we will work together to develop interim milestones.
(5) We will regularly keep the Data Model Steering Committee appraised of our progress.
(6) Our initial dialogue will be electronic, in the form of email and a web-conference site devoted to science-language issues (http://geology.usgs.gov/dm/terms/).
(7) Please access the web-conference site and register. The site was constructed and is maintained by Peter Sch-

weitzer of the USGS, who assures me that it will work as advertised (all who register at the site are supposed to be notified by email when a new contribution is made, but if anything can go wrong, it must go wrong!).

(8) My role is that of a facilitator. My job is to stimulate *your* creativity and *your* analytical approach to our task. If I am not doing that, I am failing and you must so inform me.

(9) Attached to this mail are several .pdf files, one being an archival copy of this email. My hunch is that .pdf exchange will be a common tool for the SLTT's business, so if you do not have a .pdf reader or if somehow my files are not readable by you, we need to find a fix. Please advise.

(10) The attached files include the SLTT charter, a roster of SLTT participants, and a guidelines document that the Data Model Steering Committee has reviewed, revised, and endorsed. The guidelines set the philosophy and tone of how the SLTT should go about its business. We will not be scrutinized by the DMSC, but we do have a specific mission that body expects us to achieve.

(11) I encourage you to reach out to colleagues in your organization to discuss science-language issues. We represent our colleagues and speak for them—not in place of them.

(12) Travel and travel costs: Yes, there will be a face-to-face meeting among us. As a Team, and in conjunction with the DMSC, we will have to work out the mechanics of a face-to-face, and how (or if) funds can be identified to defray (not subsidize) travel costs. I think a face-to-face is essential, for it will unite us in our task and because interpersonal exchange of ideas is always better than the impersonal electronic forum. However, travel has its costs (time and money), and such costs cannot be treated in cavalier fashion. We will discuss this as we go along.

(13) Finally, if you have searched your gut and truly do not want to participate in the SLTT process, or have second thoughts owing to the press of other obligations, please inform me as soon as possible. I will wish you well and find a replacement for you. It is essential that we all feel good about this process, and truly want to be a part of it.

I will be on the road for most of this week, and not able to check my email for much of that time. Please use this period before 17 April to get yourself into the swing of things regarding geologic-map standards. I will be back in contact next week with more mechanical issues.

In the meantime, here is our first task:

In order to set the tone for our task and see how each of us views the information content of a geologic map, please come up with 20 data-base queries that you personally would want to launch at a digital geologic-map data base. We can exchange these query-lists by email, and post them at the web-conference site. Use the following syntax:

(1) show me all metasedimentary rocks;
(2) show me Paleozoic and Late Proterozoic metasedimentary rocks intruded by Cretaceous 2-mica monzogranite;
(3) show me all low-angle faults, irrespective of their extensional or contractional origin;
(4) show me all rock units affected by two generations of folding;
(5) show me all slope-failure deposits;
(6) show me all slope-failure deposits of slump-block and earth-flow origin;
(7) show me all surficial deposits with well-developed Bt soil horizons.

I will come through with my 20, but this quick sample represents just a smat
tering of topics and issues that I would need to retrieve from a typical geologic-map data base in southern California. Good luck, and have fun.

Personally, I am looking forward to working with all of you. Collectively, we represent a considerable body of common sense, scientific breadth, and geologic-map experience (either on the data-production side or the data-use side). With such a mix, I am confident that we will do justice to the notion of common standards for geologic-map terminology—or, if such standards can not be developed and adopted, then at least a good set of minds will have reached that conclusion.

Adios from Tucson, Jonathan

MEMORANDUM FROM SLTT CHAIR TO SLTT COMMITTEE MEMBERS (4/03/2000)

SCIENCE-LANGUAGE TECHNICAL TEAM

Guidance from the North American data-model Steering Committee
04/03/2000

MANDATE

The Science Language Technical Team (SLTT) is mandated to develop standardized nomenclature for digital geologic-map data bases, including (but not limited to) the following areas:

(1) nomenclature for the description and characterization of geologic-map units (lithology, stratigraphy, geomorphology, pedology, petrology, genesis, etc.)
(2) nomenclature for the description and characterization of linear geologic features (contacts, faults, fold axial traces, mapped marker units, geomorphic features, etc.),
(3) nomenclature for the description and characterization of point geologic features (structural points, etc.);
(4) nomenclature for descriptive and interpretive information about spatial and geologic relations among geologic map units, linear features, and point features (e.g., sequencing relations, stratigraphic relations, and geometric relations, etc.).

GUIDING PRINCIPLES

(1) The SLTT's focus is digital geologic-map data bases—NOT geologic maps. Geologic maps as cartographic products should be viewed by the SLTT as derivative output FROM the data bases, not as mainline products supported BY the data bases;
(2) The SLTT's focus is the geoscience content of geologic-map data bases—not data-base design. SLTT recommendations and decisions regarding geoscience concepts and their attendant vocabulary and inter-relations will be passed upward to the Steering Committee and laterally to the Data-model Design Technical Team for evaluation and incorporation into data-model modification and tool development;
(3) Geoscience classification and nomenclature scheme(s) should be scale-independent;
(4) Classification and nomenclature scheme(s) should allow the data-base author to describe and interpret geologic elements as richly or poorly as the data allow—even within a single data base. To support this functionality, nomenclatural items should be related hierarchically in a way that allows geologic materials and geologic structures to be described and interpreted in progressively more detail and richness while still allowing them to be grouped into progressively broader categories;
(5) Classification and nomenclatural scheme(s) should be robust enough to provide stability and consistency of usage, but flexible enough to accommodate differences owing to regional or institutional mapping traditions or mission requirements;
(6) Classification and nomenclatural scheme(s) should allow the data bases to be queried for standardized geoscience concepts and geoscience attributes—ranging from the mundane to the sophisticated. Data-base queries can be only as successful as the architecture and language of the geologic data base that is queried;
(7) Classification and nomenclatural scheme(s) should accommodate all audiences and data-base users—from the educated lay audience through the end-user in local through Federal agencies, culminating in the technical geoscience user in academic and institutional audiences;
(8) Classification and nomenclatural scheme(s) should integrate seamlessly with a broad range of interdisciplinary data bases—including (but not limited to) engineering, geophysical, geochemical, hydrologic, environmental, and geographic data bases and interactive applications.

MEMORANDUM FROM SLTT CHAIR (MATTI) TO SLTT COMMITTEE MEMBERS (12/01/2000)

SCIENCE-LANGUAGE TECHNICAL TEAM

Action Plan
1 December, 2000

SLTT colleagues:

About 15 of us got together the morning of 13 November [at Geological Society of America Annul Meeting, Reno, Nevada, 2000] to discuss general issues and to develop an action plan for our science-language activities. This document summarizes the discussions, and provides the guidance for our activities over the next few months.

Participants

Lucy Edwards (USGS)
Bruce Johnson (USGS)
Ron Kistler (USGS)
Alison Klingbyle (GSC)
Diane Lane (USGS)
Steve Ludington (USGS)
Jim MacDonald (Ohio Geological Survey)
Jon Matti (USGS)
David Miller (USGS)
Steve Richard (Arizona Geological Survey)
Peter Schweitzer (USGS)
Loudon Stanford (Idaho Geological Survey)
Andy Rorick (U.S. Forest Service)
Richard Watson (U.S. Bureau of Land Management)
Jerry Weisenfluh (Kentucky Geological Survey)

(1) What we need to do

- develop lists of **control-words** for the description and naming of geologic materials and geologic structures. Control-words are rigidly defined words whose definitions cannot be violated (sandstone has exactly one definition; monzogranite has exactly one definition; thick-bedded has only one definition);
- provide formal definition of each control-word (sources: AGI dictionary of geoscience, IUGS plutonic-rock classification, widely-cited geoscience textbooks, etc.)
- develop hierarchical classification of control-words (parent-child relationships using software to be announced) (e.g., Visio2000pro)
- provide all documentation by 30 April, 2001, including:
    (1) definitions of control-terms
    (2) diagrams of parent-child relations
    (3) Minimal boiler-plate that describes our results and places them in the context of the proposed North American geologic-map data model
- Consider developing a thesaurus approach to control-terms and their non-controlled equivalents (synonyms, related terms, proxies for control-terms)

(2) Specific components of 1.0 strawman

- For the following categories, develop control-terms for the deepest level possible in each hierarchy:
    (1) **rock name** (e.g., limestone, monzogranite, blueschist, colluvium)
    (2) **lithologic attribute** (e.g., coarse-grained, fissil-weathering, thin-bedded, unconsolidated, texturally massive, porphyritic, porphyroclastic, mullion)
    (3) **rock genesis** (e.g., marine, nonmarine, alluvial, plutonic, volcanic, fluvial, colluvial, dynamothermal, high-strain)

(4) **genetic structures** (e.g., flow foliation, eutaxitic fabric, cumulate layering, graded bedding, sole structures, slaty cleavage, earth flow,)

- If possible, develop as part of each hierarchy generic field terms that allow for general-purpose classification of materials and structures (e.g., "granitic", "basaltic", "conglomeratic", "marble", "mudrock", "cross-bedded", "gneissic" "mylonitic", "silty") so that reconnaissance observations can be recorded meaningfully in the data model
- Identify internationally-recognized geologic-time classifications that can be used by the data model. The SLTT does not have to recommend or advocate any one scheme: we merely have to collect them together as schemes that can be used by the data producer. The data model design team will develop a metadata technique for associating an age term with its time-scale scheme. Time scales that come to mind include:
  (A) Harland and others (1989)
  (B) IUGS timescale (Remane, 2000)
  (C) time scales compiled in Berggren and others (1995)

(3) Target Audience

Science language should be technical—that is, it should be developed by and speak to the trained geologist. Although we all are concerned about how the professional and non-professional non-geoscience audience will access and understand our database content, this concern should be addressed by a technical team tasked with designing the data-model user interface.

(4) Basis and scale of terminology

Map-unit categories (i.e., formation, member, tongue, lentil, bed) are conceived and extended through a process that integrates hierarchical observations beginning at the ***hand-sample and outcrop level*** but extending to the ***hillside and regional level*** and augmented by the ***thin-section and chemical-analysis*** level. Thus, hierarchical terminology schemes leading to map-unit description should reflect:

- regional-scale observation
- hillside-scale observation
- outcrop-scale observation
- hand-sample-scale observation
- thin-section-scale observation
- chemical analysis-scale observation

(5) Existing strawman-classifications for consideration include (but are not limited to):

- Rock classification schemes of British Geological Survey (BGS)
- Version 6.0 classification scheme of SLTT member Bruce Johnson (Matti will distribute again; Johnson will provide parent-child diagrams)
- Volcanic and plutonic classification schemes of SLTT member Steve Ludington (Matti will distribute again)
- SCAMP version 2.0 rock-classification schemes (Matti will distribute again)
- Any other hierarchical classification schemes that subgroup members can identify

(6) In addition to nomenclature for sedimentary, igneous, metamorphic, and surficial materials, we need to develop language for the following materials:

- tectonic rock units (e.g., broken formations, mélanges, tectonic breccia, bolide-impact rocks)
- rock-types of hydrothermal or alteration origin
- rock-types of mixed origin
- rock-types of unknown origin

(7) The following rules MUST be adhered to:

- hierarchies must follow independent non-intersecting pathways (or so I understand [correctly?] from the data model design people)
- A control-term cannot be arrived at by more than one pathway. For example, the mineral "calcite" cannot be arrived at via a sedimentary pathway leading to calcite or a metamorphic pathway leading to calcite or an igneous pathway leading to calcite. Instead, the mineral calcite must be approached via a single pathway in a mineralogy hierarchy that incorporates children of calcite (e.g., calcite, *sedimentary*; calcite, *metamorphic*; calcite, *vein*).

(8) To assist data-model design team, we need to distinguish between the following terms:

- "rock"
- "rock unit"
- "map unit"

(9) To assist data-model design team in developing a map-unit characterization field

- develop language that allows each map unit to be characterized concisely and distinguished clearly from other map units
- develop control terms applicable to lower, upper, and lateral **boundaries of map units** (e.g., conformable, unconformable, sharp, discrete, transitional, gradational, mixed, migmatitic, intrusive, extrusive, interfingering) and for distinguishing properties (geologic, geomorphic, pedogenic, paleontologic. This may not be possible within the scope of our initial lithologic assignment, but we need to have it on our radar screen as we do our job, and make some progress in this direction.

References Cited

Berggren, W.A., Kent, D.V., Aubry, M-P., and Hardenbol, J., eds., 1995, Geochronology, time scales and global stratigraphic correlation: Tulsa, Oklahoma, Society of Economic Paleontologists and Mineralogists, Special Publication 54, 386 p.

Harland, B.W., Armstrong, R.L., Cox, A.V., Craig, L.E., Smith. A.G., and Smith, D.G., 1989, A geologic time scale, 1989: New York, Cambridge University Press, 246 p.

Remane, Jurgen, compiler, 2000, International Stratigraphic Chart: Division of Earth Sciences, UNESCO, 16 p., 1 plate. Published in cooperation with the International Union of Geological Sciences and the International Commission on Stratigraphy.

MEMORANDUM FROM SLTT CHAIR (MATTI) TO SLTT SEDIMENTARY AND SURFICIAL SUBGROUP MEMBERS (2/15/2001)

Sedimentary and Surficial Subgroup Colleagues: 02/15/2001

Now that the surficial and sedimentary SLTT teams both have launched their deliberations, we need to address an issue of concern to both teams.

Several members of the sedimentary group have indicated concern about possible overlap between "unconsolidated" sediment and unconsolidated surficial materials. This concern originates from the British Geological Survey's (BGS) classification of sedimentary materials into "lithified" and "unlithified" materials:

"The primary classification of sediments and sedimentary rocks is based on their compositional attributes present at the time of deposition. This allows sediments to be classified by the same compositional boundaries as sedimentary rocks."

The following points address this issue, and seek to clarify the unique assignments of the surficial-materials team and the sedimentary team. As you see fit, please comment on any of the points:

(1) I do not think the BGS blurs the boundaries between their "unlithified sediment" and "unconsolidated surficial materials";

(2) The BGS scheme simply provides names for unlithified sediment that are parallel with the names for lithified sedimentary rock;

(3) The surficial team is charged with developing a classification of surficial materials like "alluvium", "colluvium", "landslides", and so forth. I suspect that the group will come up with classification categories such as "alluvial deposits", "colluvial deposits", "landslide deposits", etc., all representative of surficial materials that are relatively "unlithified";

(4) These surficial materials will have certain physical properties (such as grain size, particle shape, bedding thickness, grain composition, grain-matrix ratios, color, etc.) that will overlap with the physical properties of sedimentary materials, both lithified and unlithified. This overlap will be especially obvious between surficial materials that are water-laid and unlithified sediment that is water-laid: the two are one and the same, are they not? And that is the source of the apparent overlap;

(5) Should both the surficial team and the sedimentary team independently create classification schemes for (a) unlithified sand bodies that form bars on the Platte braided-river plain or (b) unlithified sand bodies in coastal chenier plains or (c) unlithified oolitic shoals in the Bahamas or (d) unlithified mudrock and channelized sand bodies in the Mississippi River delta?

(6) My answer to question (5) is "no". I expect that the surficial team will view those specific examples as surficial materials that could be classified and named and mapped as (a) alluvium, braid-plain type or (b) paralic deposits, chenier-plain type or (c) marine surficial deposits, carbonate, oolitic-shoal type or (d) alluvial deposits, deltaic (to name some hypothetical possibilities);

(7) I believe that the physical properties of unlithified deposits and the *naming of specific sediment types they contain*, are the purview of the sedimentary team. This is the position the BGS takes, I believe;

(8) Thus, for points (5) and (6): (a) alluvium, braid-plain type, may consist of medium-bedded, texturally massive to flat-laminated, moderately sorted, medium- to coarse-grained quartzofeldspathic lithic sand, while (b) paralic deposits may consist of crudely bedded flat-laminated to trough-laminated, well-sorted, fine- to medium-grained quartz arenite sand while (c) marine surficial deposits, carbonate, oolitic-shoal type may consist of.......etc.;

(9) I think the tasks of the two teams will be clarified if we adopt the following:

- the surficial team is classifying deposit types that can be used as map units, and that also may occur within map units, but are not specific lithologies or petrographic sediment names
- the sedimentary team is classifying rock types that occur as specific lithologies in outcrops and in map units
- the sedimentary team will develop much of the classification and description nomenclature for specific rock types, but they must do so in partnership with the surficial team so that cross-pollination occurs

(10) The distinction between "consolidated" and "unconsolidated" or between "lithified" and "unlithified" is going to be a vexing issue, irrespective of the issues raised in the preceding nine points. I will not venture into this now.

Please ruminate over the ideas in this note. If we are not on the same page on this one, we could get into trouble. I am just thinking out loud, so give it your own treatment.

Adios, Jonathan

MEMORANDUM FROM SLTT CHAIR (MATTI) TO SLTT COMMITTEE MEMBERS (3/14/2001)

SLTT colleagues: 03/14/2001

Recent discussion within some of the subgroups indicates the need to restate and clarify the purpose of our SLTT goals and the nature of our classification activities. Please read this memo carefully and at your earliest convenience. If any of you has major reservations, concerns, or disagreements about these objectives, please raise them to all of us now.

By separate mailing, I am sending a copy of this memo to the North American Data Model Steering Committee, whose members also are asked to comment and evaluate the statements.

(1) Databases versus geologic-maps: *Our purpose is to develop classification structures for digital geologic-map databases*, not for digital versions of geologic maps. The production of a geologic-map plot is incidental to the database, and is not the primary focus of the language that the SLTT is developing.

The difference in tone here is important: the hierarchical structure, number of rock classes, and other aspects of our language schema should be tailored to storing and searching science concepts in a digital database, not tailored to the requirements of database fields in a particular data model or tailored to the text in a geologic-map legend.

(2) Language for new data *versus* language for compilation: While the compilation of pre-existing geologic mapping obviously is part of a geologist's activities, the SLTT's primary driver is to develop schema that *facilitate the classification and communication of new field information*. We must look into the future toward novel ways of organizing new data, not into the past to find ways of facilitating the compilation of old data. The former will benefit the latter in obvious ways.

Map compilation (the collation, evaluation, interpretation, and translation of geologic-map information contained in products produced by other workers) is a necessary and legitimate goal. However, the creation of science language that supports geologic-map compilation is not the SLTT purpose.

(3) Do we need to accommodate pre-existing science language?: Compilation of pre-existing geologic-map information requires the geologist to deal with a wide array of lithologic names and descriptors that have come down through the generations. Should the SLTT classification schema create a place for these terms, or define equivalencies for them?

No. *Our task is to create a single uniform, coherent classification that logically, objectively, and thoughtfully establishes rock names and descriptors that classify geologic materials accurately and comprehensively according to modern usage*. We are not obliged to create a list of synonyms or equivalencies. We are not necessarily required to make a place for previous usage, no matter how entrenched that usage might be.

For the compilation of pre-existing map information, it is (and always should be) the responsibility of the map compiler to interpret what a published geologic map contains, and to place this information in the context of modern rock classifications. This is why geologists (who have the training and expertise to make geologic judgments) should be map compilers. The SLTT classification schema will be *the* modern standard for geologic-map database attributes. *It will be the responsibility of future map compilers to interpret the nomenclature of pre-existing geologic-map information for its position in the SLTT schema, not the responsibility of the SLTT to accommodate all previous language*. Pre-existing language should be treated either in feature-level metadata or dataset metadata: this will create a paper trail for original usage, but will not burden the SLTT schema with the diverse nomenclature of the past.

(4) Language for data producer *versus* end user: *The lithologic classification schema we are developing are NOT for the end user, but for the geologist who is collecting attribute data and populating a database with the attributes*. The production of derivative databases and map plots that serve end users is not the SLTT concern.

Does this mean that the SLTT is not mindful of end-users? Nope. Each of the four subgroups is working hard to develop science language that will form a foundation for users of all kinds—from technical to non-technical. But the SLTT focus needs to be geologist-directed in order for the multiple-user base to be served.

We all are interested in and concerned about how end users access and use geologic-map information. However, I strongly believe that the proper focus of end-user facilitation should be the design of an appropriate user-interface. It will be the job of (a) the SLTT, (b) the data-model design team, and (c) a user-interface team (currently not designated) to design an appropriate tool-set to take the concepts and language designed by the SLTT and make them user-friendly.

(5) Hand-sample language *versus* map-unit language: *The SLTT mandate is to provide classification schema for individual rock types that occur in geologic-map units, together with language that describes the physical appearance, composition, and genesis of these rock types*. The science language must focus on hand-sample and outcrop-scale attributes, but should include rock names and fabric relations that derive from thin-section observations as well as language for sequencing and stratigraphic relations at the map-unit scale.

(6) How comprehensive or finite should our classification schema be? *Our science language should reflect the reali-*

ties of geology, not the requirements of end users. However, the geologic universe is complex, so should the classification schema be complex and opaque? Nope. And that is the challenge: to represent rock names and rock structures within families that bring order to the complexity.

In a note to the metamorphic subgroup, Bruce Johnson correctly pointed out that " if the classification is hierarchical, and the first and second levels of the hierarchy are limited to a small number of classes, then it becomes possible to render the map by ignoring lower levels". Bruce's concern here is that the plethora of detail that we could create in our classification schema should not bar the database user from perceiving the major high-level relationships among geologic elements. I agree completely. However, if logically structured, then the number of classes or branches or levels of the hierarchy will not matter.

In my opinion, the user interface will be THE critical device for sorting through the database from higher (general) levels to lower (detailed) levels to accommodate user needs.

(7) Do we need flow charts and glossaries?: To the extent that we must define control terms and root names, etc., then to that extent we are defining a glossary of terms. One strength of the British Geological Survey Rock Classification schemes is the decision-making pathways (flowcharts) that the schemes establish for the use of data producer and data-compiler (and ultimately, from an interface point of view, the data user). A decision-support mechanism is a natural fallout of control terms: the terms must have definitions, and a decision process must be executed in order for a control term to be used or not used by the geologist and end-user. A flow chart is a logical device for displaying the decision-support process.

Let me end by sharing what I am discovering while working with the sedimentary subgroup. In my opinion, we need hierarchical classification schema that allow the geologist to go as deep into the data-attribution process as possible—without getting painted into a corner. The BGS sedimentary classification scheme doesn't have a lot of wiggle room in it. For example, feldspar-rich sedimentary rocks are termed "feldspathic arenites" as defined by Pettijohn. End of statement. End of choices. I personally would be more comfortable if an intermediate level existed that gave the geologist (and the end user) more generic terms like "feldspar-rich" or "lithic-rich", *before* requiring the geologist (and the end user) to commit to the name "feldspathic arenite". This would allow me to classify a rock in the field as a "feldspar-rich sandstone" based on hand-lens observation, and I could stick with this name if I never obtained modal data that would allow me to document the rock as a feldspathic arenite (*sensu* Pettijohn). My audience can get a lot out of the term "feldspar-rich sandstone", even though I haven't tagged the rock as a "feldspathic arenite" *sensu strictu.*

In other words, common sense needs to drive our process—and I ask that you work with each other to find this common sense. A purely academic approach to rock classification and description is not going to do us any good. Even though I minimized the role of the end-user as a target for our deliberations, none-the-less both the field geologist and the land-use manager need a classification that allows each to (a) classify a rock in as much detail as desirable and (b) search the forest before searching out individual trees.

In other words, we do not have an easy job.

Adios, Jonathan

Conversion of Lithological Data in the Manitoba Water Well Database (GWDrill) to a Mappable Format

By L. H. Thorleifson[1] and D. M. Pyne[2]

[1]Geological Survey of Canada
Current address: Minnesota Geological Survey
2642 University Ave W
St Paul, MN 55114-1057
Telephone: 612-627-4780 ext 224, Fax: 612-627-4778
e-mail: thorleif@umn.edu

[2]Geological Survey of Canada
601 Booth Street
Ottawa ON K1A 0E8

ABSTRACT

A project designed to gain access to all usable data to aid construction of 3D geological models for southern Manitoba has included significant effort to appropriately utilize lithological data of variable quality in the Provincial water well database known as "GWDrill". In order to use the data, it was necessary first to assign x, y, and z coordinates to each site on the basis of the existing designation, usually a quarter section (one-half mile by one-half mile) or river lot location. Secondly, it was necessary to convert the lithological data to a classification and terminology that could be queried and mapped. The lithological data were converted by correcting spelling, obtaining an inventory of words, deleting unusable words, identifying synonyms and changing them to a single term, and parsing and interpreting the remaining information into several lithological, colour, structure, consistency, hydrogeological, and stratigraphic variables. The resulting database has exceeded expectations with respect to apparent location accuracy and geological coherence.

INTRODUCTION

As a result of the foresight of Manitoba Water Resources Branch staff, the Province of Manitoba has an excellent digital database of water well records, known as GWDrill. The database includes several tables, including a table of verbal stratum descriptions. These lithological data typically appear unmodified from the driller's reports, providing very useful site-specific information in the driller's original wording.

Progress in application of computer-based mapping methods has, however, increased recognition that data would ideally also be formatted for use in these applications. In doing so, caution is required because data originally reported with a low level of certainty may be formatted for computer-based applications in a manner that implies a higher level of certainty than was originally intended. One approach to this challenge is to set up procedures that allow many drillholes to be viewed at once, allowing anomalous data to become more apparent. Doing so requires that sites be assigned x, y, and z coordinates so that sites can be graphically portrayed relative to each other, and that lithological data of varying format and terminology be translated to a set of attributes under which consistent, although evolving, terminology is used.

This paper describes an experiment in the implementation of this approach to the Manitoba water well database. The objective was to permit lithological data to be queried and mapped in 3D, thus aiding the construction of geological models for southern Manitoba.

LOCATION AND ELEVATION

Virtually all of the records in the database are located to a legal survey polygon such as a quarter section or a river lot, so a database of x and y locations for centroids of the legal survey polygons was required. Two formats are present: Dominion Land Survey (DLS) quarter sections, and Parishes, consisting of river lots and other survey systems that predate the DLS grid. Legal survey data obtained from Manitoba Conservation were used to obtain x and y values for the centroid of each polygon. Sites located to quarter sections were assigned the quarter section coordinates. Wells located to a section (1 mile by 1 mile) were assigned a location at the centroid of the section, obtained by averaging the quarter section xy values. Wells assigned locations no more specifically than section, for example to a township (6 miles by 6 miles), were not assigned coordinates for estimated location. Wells assigned to a river lot were given coordinates for the centroid of the lot. Where river lots are divided into two or more portions, such as inner and outer lots divided by a road, the average of the coordinates was used for a well whose legal survey location did not indicate which portion of the lot the well was located in.

Elevations were assigned by intersecting the legal survey polygon centroid x and y values with a digital elevation model derived from legal survey data that typically is accurate to +/- 3m.

PREPARATION OF THE DATA

Processing was first applied to a November 1998 version of the database that consisted of 83,597 sites (fig. 1) and 402,461 strata. Subsequently, the database was updated with a July 2000 version consisting of 87,992 sites and 422,917 strata. Ten sites that had occurred in the 1998 version had been deleted prior to release of the 2000 version, and 35 strata in the 1998 version do not appear in the 2000 version. Using both versions, a database of 88,002 sites and 422,943 strata resulted.

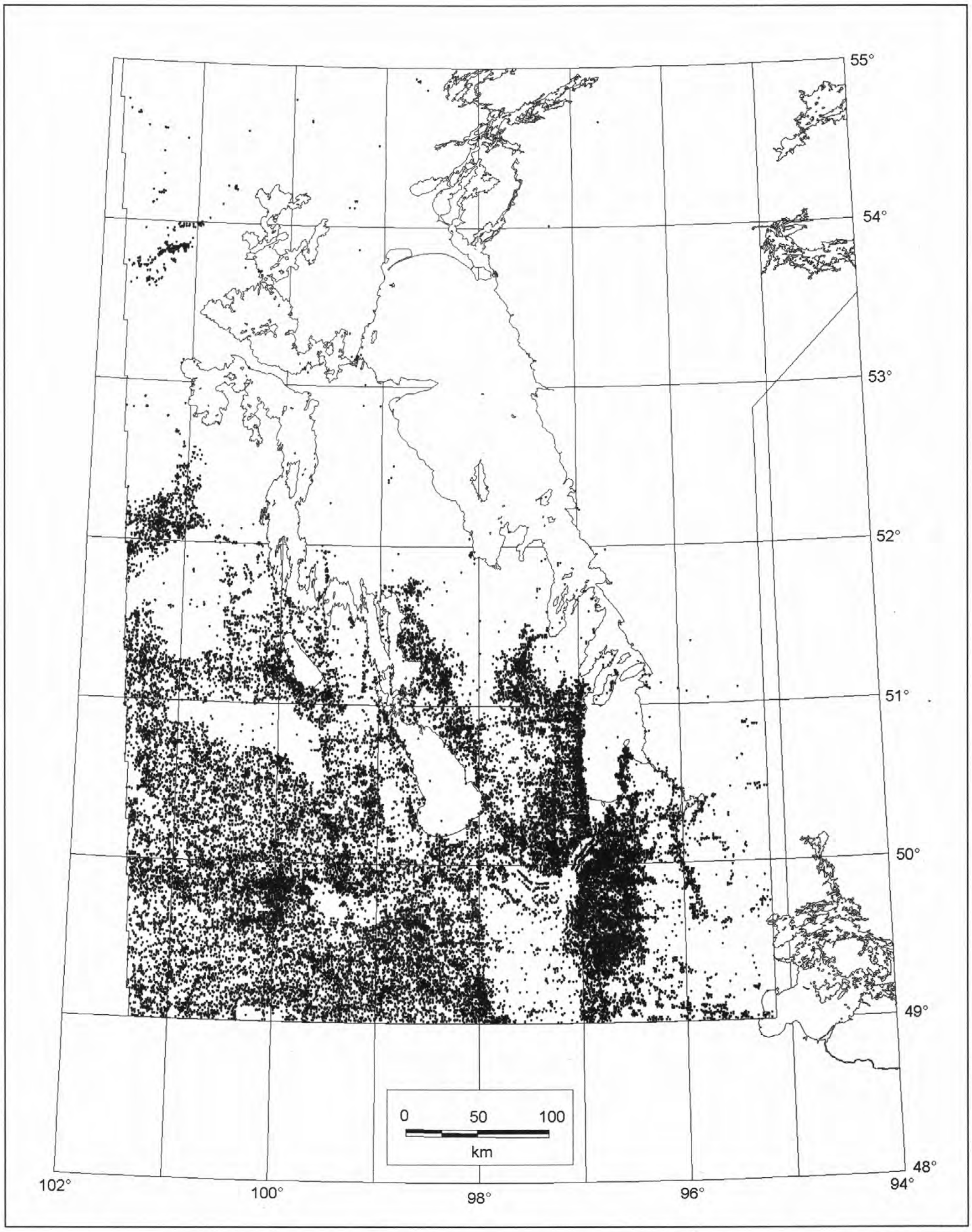


Figure 1. Location of GWDrill sites south of 55°N; 166 additional sites occur farther north.

The database download in 1998 was acquired from Manitoba Water Resources Branch twice, due to media format problems. The first version used memo fields of unrestricted length for the lithologic description, whereas the subsequent version used 255 character text fields, among which 15 records were truncated. The truncated records were restored after a method was found to read the first download. In addition, the first version contained both UTM zones 14 and 15, whereas the second, which was implemented, exclusively used zone 14.

It was agreed by all involved that no changes would be made to the driller's original lithological description as it appeared in the database, and Geological Survey of Canada (GSC) staff have no access to the permanent data archive. The intention was to generate a table of parsed information that could be linked to the original stratum file, with no change in the number of strata.

In addition to this table, however, an add-on table of zero-thickness strata was created to capture information regarding the nature of material at a lower contact. This information typically takes the form of a statement such as: 20 to 30' sand, rock at 30'. The new zero-thickness stratum was given an incremented sequence number, and the original description was modified in the add-on table, for example, sequence x, 20-30' sand, sequence x+1, 30-30', rock. These data were considered important for mapping the bedrock surface, as well as the top of Precambrian rock where it underlies Phanerozoic cover.

Records for the 88,002 sites include descriptions for 422,943 strata. An average of about 3 words is used per stratum, hence a total of about 1.2 million words occur in the original data. These words are arranged in 77,457 different combinations. Conversion of the lithological data commenced by correcting spelling errors using a word-processor spell-check of a copy of the descriptions. Having done so, the number of unique stratum descriptions was reduced to a total of 72,467. These records are constructed from 1,956 different words, excluding numbers and punctuation.

A table of these words was obtained by inserting a carriage return in a copy of the descriptions after all words, and by using database procedures to obtain a word inventory with frequency of occurrence. Each of the 1,956 words was classified with respect to whether the word was to be deleted, substituted with a synonym, or retained, and all words were classified with respect to a list of topics that emerged as the analysis progressed. These included lithological, colour, structure, consistency, hydrogeological, and stratigraphic variables.

Over half of the words were deleted, including terms such as 'and', 'with', and 'of' (table 1) that were of use in a phrase, but are not essential in parsed information. It should be noted that several thousand strata were re-interpreted in the late stages of processing on the basis of searches and inspection of the original description, and at this stage, the word 'and' was a factor. Hence some words that were deleted for the first iteration may later have been considered during a final iteration of interpretation.

Over one-third of the words were synonyms, and were changed to common terms. Among terminology related to lithology (table 2), the words most commonly replaced were synonyms of 'gravel', including 21 different words. Other words that occurred in many synonym forms are 'previous well', 'fill', and 'fossiliferous'. Some synonym replacements are at equal levels of specificity, such as 'hardpan' and 'till', while other substitutions are from differing levels, such as 'fossiliferous' and 'stromatolitic'.

Many changes were made to terminology related to colour (table 3), in order to convert all colours to Munsell word equivalents. The colour to which the greatest number of changes was made was 'light yellowish brown', including the commonly-occurring words 'buff' and 'beige'.

In the case of words related to structure and consistency (table 4), the words most commonly replaced were synonyms of 'stratified', including 55 different words. Other words that occurred in many synonym forms, or in some cases subordinate classes, are 'hard', 'soft' and 'firm'. Among words related to hydrogeology (table 5), the term with the largest number of synonyms was 'fractured/permeable', including 37 different words. Other words that occurred in many synonym forms, or in some cases subordinate classes, are the term 'saturated' and terms related to effervescence or odour.

A total of 92 words was retained from the original description without modification (table 6). In addition, numerous words related to stratigraphic nomenclature were retained (table 7).

Following consultations with co-operating agencies regarding the scheme, the deletion and synonym substitution procedures were applied to a copy of the spell-checked lithological data. Multiple copies of the result were made, and each copy was assigned to a topic such as lithology, colour, structure, or hydrogeology. In each copy, only words considered relevant to that topic were retained. This was done by substituting an ASCII symbol for each of the eligible words, deleting all letters of the alphabet, and then changing the symbols back to the desired words. The retained words were then either parsed into multiple fields, or interpreted. During this procedure, it often was necessary to refer to the original record and apply judgement based on geological knowledge, including the context of the stratum in a sequence. Location was not, however, considered.

CLASSIFICATION AND TERMINOLOGY

In order to generate a database that could readily be queried and mapped, it was essential to parse the information into a set of variables, under which a minimum number of common terms with no synonyms is used. Synonyms were not allowed, so that a query can be carried out without having to list the multiple words that may

Table 1. Most commonly occurring words that were deleted, with frequency of occurrence.

AND	64758
WITH	15199
OF	11704
FEET	10715
AT	8405
SOME	5240
THIN	1688
SMALL	1684
ODD	1187
IN	1080
OR	946
FEW	785
SLIGHTLY	762
PIT	733
LITTLE	608
A	521
NO	500
FAIRLY	498
TRACES	480
GRAINED	468
ROUGH	428
OVERLAY	403
LEVEL	399
GLACIAL	398
MORE	373
COLOURED	362
MOSTLY	362
DEPTH	362
COARSER	345
BOTTOM	323
LEDGES	306
LARGE	274
ABOVE	266
ALLUVIAL	242
LOTS	233
NOT	230
FROM	228
INCH	207
UNIFORM	205
DRILLING	195
AS	189
HEAVY	189
LIKE	179
RIVER	173
MAINLY	171
CAVING	163
DUG	162
PURE	162
HIGHLY	152
SMOOTH	146
AFTER	144
LAKE	140
THAN	137
ZONE	136
GRAINS	134
GOOD	132
THROUGHOUT	132
PACKED	129
OCCASIONAL	127
MATERIAL	124
MINOR	120
BASE	119
HOLE	115
OPEN	115
LACUSTRINE	113
COLOUR	107
BACK	105
NUMEROUS	105
TOP	102
MUCH	102
ON	100
BELOW	98
GPM	96
NICE	95
BECOMING	95
BUT	94
IS	94
LOOSELY	92
BIT	92
INCHES	91
CHANGING	90
MATRIX	90
FINER	88
DRILL	86
LESS	85
LOST	85
IGPM	84
LOSS	82
DRILLED	82
TAKING	81
SAME	80
HIGH	79
MODERATELY	79
AMOUNT	74
ALMOST	74
IT	71
REPORTED	71
TURNING	70
SIZE	70
LOOKS	69
PROBABLY	69
GETTING	68
DARKER	67
BIG	66
SEMI	65
POORLY	63
UNKNOWN	63
SURFACE	63
PLATES	63
LOSING	63
PREDOMINATE	63
THE	62
POSSIBLY	61
ETC.	61
SHARP	59
DIAMETER	59
OFF	58
BETWEEN	58
DECAY	58
NEAR	57

Table 2. Lithology-related synonyms that were changed to common terms, with frequency of occurrence in parentheses.

GRAVEL	11593	stones (5818), pebbles (3249), cobbles (597), fragments (538), pieces (275), particles (269), chips (267), bits (244), cobble (135), pebble (70), granular (34), cobblestones (24), granules (23), cobblestone (16), gravels (11), clasts (7), granulars (7), chip (5), chippings (2), aggregate (1), gravel-cobbles (1)
TILL/DIAMICT	10328	hardpan (10213), till-like (62), tilly (41), colluvium (3), semihardpan (3), tillish (3), tills (2), claybound (1)
SOIL	6630	topsoil (5289), loam (934), sod (274), dirt (85), earth (15), roots (12), vegetation (11), manure (5), turf (3), mould (1), soils (1)
GRAVELLY	6208	stoney (5607), pebbley (577), cobbley (18), stonier (6)
UNDIFFERENTIATED ROCK	4332	bedrock (2531), stone (1748), caprock (51), sedimentary (2)
INTRUSIVE OR HIGH GRADE METAMORPHIC	3789	granite (3343), granitic (333), igneous (90), schist (9), gneiss (8), pegmatite (4), diorite(2)
PREVIOUS-WELL	2369	well (1259), old (667), existing (227), casing (31), previously (29), crib (28), cribbed (26), excavated (24), abandoned (23), cribbing (16), cased (15), previous (8), pipe (8), deepened (4), trench (3), wells (1)
BOULDER(S)	1749	rocks (1743), block (3), boulders-cobbles (2), blocks (1)
SILICA/SILICEOUS	1578	silica (889), quartz (621), siliceous (54), agate (7), agates (3), tripolized (1), quartzose (1), silica-like (1), chalcedony (1)
BENTONITIC SHALE/MUDSTONE	1052	bentonite(1052)
FILL	889	road (364), backfill (291), roadbed (136), concrete (28), ash (23), pavement (9), asphalt (8), roadfill (6), driveway (5), rockfill (3), culvert (3), garbage (2), backfilled (2), sawdust (2), landfill (1), fill-type (1), runway (1), bricks (1), roadway (1), junk (1), dike (1)
EVAPORITES	788	gypsum (752), gyprock (21), anhydrite (15)
CARBONATE	783	lime (763), tyndall (8), carb (6), carbonates (5), calcarenite (1)
DOLOSTONE	756	dolomite(737),dolostone(19)
SAND	549	quicksand (482), beach (22), sands (21), grit (21), alluvium (2), sand-like (1)
LOW GRADE METASEDIMENTARY OR METAVOLCANIC	431	soapstone (223), slate (118), volcanic (33), greenstone (18), quartzite (11), marble (9), basalt (9), metamorphic (7), andesite (1), argillite (1), volcanics (1)
CLEAN	400	washed (249), cleaner (149), cleaned (1), cleanest (1)
FINE-TO MEDIUM	360	fine-medium (253), medium-fine (106), fine-medium-fine(1)
UNCLASSIFIED	297	log (297) *(inherited from 'no log' after 'no' was deleted)*
FOSSILIFEROUS	264	shells (163), shell (51), clam (12), fossils (12), fossil (7), snail (7), bone (3), crinoid (2), brachiopods (2), bones (1), mollusk (1), snails (1), stromatolitic (1), vertebrae (1)
PEAT/ORGANIC	258	muskeg (85), swamp (75), muck (33), moss (30), quagmire (17), swampy (16), peatmoss (2)
SULPHIDE-BEARING	257	pyrite (228), pyritic (12), pyrites (5), sulfide (5), marcasite (2), arsenopyrite (2), sulfides (1), ore (1), metallic (1)
MEDIUM-TO COARSE	238	medium-coarse (232), coarse-medium(6)
SOFT CLAY	236	gumbo(183), putty(53)
FINE-TO COARSE	220	fine-coarse (215), coarse-fine (3), fine-medium-coarse (2)
SILTY CLAY	217	mud(216),silt-clay(1)
UNDIFFERENTIATED SOIL/SEDIMENT	217	overburden (168), drift (19), sediment (18), surficial (7), sediments (4), regoliths (1)
SANDY	193	gritty (184), sandier (5), loamy (4)
WOOD-BEARING	182	wood (179), woody (2), stump (1)
SHALE/MUDSTONE	181	siltstone (105), shale-like (31), shales (28), claystone (7), mudstone (4), quartz-shale (2), clay-shale (2), shalestone (1), diatomaceous (1)
BOULDERY	179	rocky (179)
CLAYEY	154	muddy (113), clayed (25), clayish (14), clay-type (1), clayier (1)
COAL	144	lignite (142), coals (2)
PETROLIFEROUS	144	oily (112), oil (27), petroleum (4), hydrocarbons (1)

Table 2. Continued.

FERRUGINOUS	137	iron (77), oxide (30), hematite (17), oxides (10), ferrous (1), sideritic (1), ironstone (1)
CONCRETIONARY/NODULAR	114	nodules (49), concretions (33), concretion (29), nodular (3)
WELL SORTED/GRADED	71	sorted (48), graded(23)
VOID	70	sinkhole (48), cavern (7), cavity (5), drop (3), cave (3), sink (1), cavernous (1), caves (1), cavities (1)
SILT	67	fines (30), silts (29), flour (8)
CALCAREOUS	51	calcite (29), limey (13), calcium (7), calcitic (1), marly (1)
CHERTY	50	chert (32), flinty (14), flint (4)
MEDIUM	48	pea (47), pea-size (1)
RUBBLE	47	rubbley (26), breccia (20), rubbles (1)
CALCAREOUS SILT	39	marl (39)
AUTHIGENIC-XL-BEARING	38	crystals (24), selenite (12), roses (1), rosettes (1)
MICACEOUS	38	mica (29), biotite (4), biotitic (3), muscovite (1), schistose (1)
SILTY	36	silted (20), siltier (14), silting (2)
CLAY	35	clays (34), soil-clay (1)
CARBONACEOUS/BITUMINOUS	20	tar (10), charcoal (3), coaly (2), bituminous (2), sooty (1), carboniferous (1), tar-like (1)
FELDSPATHIC	12	feldspar (12)
ROUNDED	11	round (9), spherical (2)
ORGANIC	10	peaty (6), plant (2), carbon (1), marshy (1)
DIRTY	9	impure (9)
KAOLINITIC	8	kaolin (5), kaolinized (3)
LIMESTONE	2	ls (1), quasi-limestone(1)
CHALKY CARBONATE	5	chalk(5)
OOLITIC	3	oolites (2), oolite (1)
SANDY SILT	3	sandy-silty (3)
CARBONATE & PRECAMBRIAN-RICH	2	granite-limestone (2)
MEDIUM-TO VERY	2	medium-very(2)
SHALEY	2	shalier (2)
SILTY SAND	2	silt-sand(2)
SUBROUNDED	2	subspherical (2)
ARGILLACEOUS	1	argillitic (1)
CARBONACEOUS/BITUMINOUS SHALE/MUDSTONE	1	coal-shale(1)
FINE-TO VERY	1	fine-very(1)
PRECAMBRIAN-RICH	1	granity (1)
SANDY CLAY	1	silt-sand-clay(1)

Table 3. Colour-related synonyms that were changed to common terms, with frequency of occurrence in parentheses.

LIGHT YELLOWISH BROWN	1708	buff (1167), beige (531), blonde (2), buff-brown (2), brownish-buff (2), buff-tan (1), tan-yellow (1), buff-grey (1), beige-grey (1),
VERY-LIGHT BROWN	361	cream (359), whitish-brown (1), cream-white (1)
BROWNISH GREY	347	brown-grey (264), brownish-grey (77), tan-grey (2), grey-brownish (2), grey-cream (1), grey-buff (1)
GREYISH BROWN	336	grey-brown (287), greyish-brown (46), grey-tan (2), brown-greyish (1)
MULTICOLOURED	327	varicoloured (90), specks (73), speckled (68), flecks (36), spots (25), patches (15), speckles (6), mottling (5), marbled (3), spotted (3), variegated (2), mottle (1), light-dark (1)
LIGHT BROWN	326	tan (279), cream-brown (29), brown-white (5), tan-brown (4), white-brown (3), light-brown (3), brown-light (1), light-buff (1), tan-coloured (1)
LIGHT	304	whitish (179), creamy (78), lighter (36), milky (9), light-medium (1), bland (1)
REDDISH	211	rusty (210), rusty-yellow (1)
GREYISH BLACK	179	grey-black (179)
REDDISH BROWN	163	rust (67), reddish-brown (54), red-brown (23), orange-buff (7), brown-red (7), buff-orange (3), rust-brown (1), rust-red (1), reddish-buff (1)
BLUISH GREY	122	blue-grey (89), bluish-grey (32), grey-blue-grey (1)
GREENISH GREY	109	greenish-grey (57), green-grey (50), greenish-grey-white (1), grey-greenish (1)
YELLOWISH BROWN	90	yellow-brown (90)
DARK BROWN	78	chocolate (69), black-brown (7), brown-blackish (2)
GREYISH WHITE	76	grey-white (74), greyish-white (2)
SPECKLED	68	salt (36), pepper (30), salt-pepper (2)
LIGHT GREY	66	white-grey (41), light-grey (13), cream-grey (8), grey-light (2), silver (2), grey-whitish (1)
BLUISH GREEN	57	blue-green (53), bluish-green (4)
PALE PURPLE	51	mauve (51)
DARK	44	blackish (43), blackened (1)
LIGHT GREENISH BROWN	36	khaki (36)
GREYISH GREEN	36	grey-green (29), greyish-green (7)
BRILLIANT	27	bright (26), vivid (1)
BROWNISH BLACK	25	brown-black (24), black-brownish (1)
REDDISH ORANGE	23	orange-brown (11), orange-red (5), reddish-orange (3), red-orange (2), red-yellow (2)
GREENISH BROWN	15	green-brown (13), greenish-brown (1), brown-greenish (1)
DARK GREY	14	black-grey (12), blackish-grey (2)
GREENISH WHITE	14	greenish-white (11), green-white (3)
BROWNISH YELLOW	13	brown-yellow (11), brownish-yellow (2)
GREYISH BLUE	11	grey-blue (11)
BROWNISH	10	brick (10)
VERY-LIGHT GREY	10	whitish-grey (10)
REDDISH GREY	10	reddish-grey (8), red-grey (2)
WHITE	9	off-white (8), white-light (1)
BLUISH BLACK	8	blue-black (7), bluish-black (1)
LIGHT RED	7	rose (7)
LIGHT GREENISH BLUE	7	turquoise (7)
YELLOWISH GREY	7	yellow-grey (6), yellowish-grey (1)
GREENISH BLUE	6	green-blue (5), blue-greenish (1)
RED	5	ochre (5)
BROWNISH BLUE	4	brown-blue (4)

Table 3. Continued.

BROWNISH GREEN	4	brown-green (4)
GREY	4	greyer (3), pearly (1)
GREENISH BLACK	4	green-black (3), greenish-black (1)
PURPLISH GREY	4	purplish-grey (2), grey-purple (1), purple-grey (1)
GREENISH YELLOW	3	greenish-yellow (3)
BRILLIANT YELLOW	3	gold (3)
LIGHT PINKISH GREY	3	pink-grey (3)
MULTICOLOURED BLUE	3	yellow-blue (3)
PALE YELLOW	3	yellow-white (3)
PALE	3	faint (3)
BROWNISH RED	3	brownish-red (3)
GREYISH RED	3	greyish-red (2), grey-red (1)
BROWN	3	drab (2), brown-brownish (1)
YELLOWISH WHITE	3	white-yellow (2), yellowish-white (1)
YELLOWISH RED	3	yellow-rusty (2), yellow-red (1)
MULTICOLOURED GREY	3	grey-white-black (2), blue-grey-green (1)
DARK BLUE	3	blue-dark (1), black-blue (1), blackish-blue (1)
MULTICOLOURED RED	3	blue-green-reddish (1), black-red-white (1), red-blue (1)
PINKISH WHITE	3	white-reddish (1), white-pink (1), white-pink (1)
MULTICOLOURED ORANGE	2	orange-grey (2)
STRONG	2	blood (2)
LIGHT GREYISH	2	silvery (2)
LIGHT PINKISH BROWN	2	pink-beige (2)
GREEN	2	greener (2)
MULTICOLOURED BROWN	2	white-brown-black (1), grey-white-brown (1)
PURPLISH RED	2	purplish-red (1), purple-red (1)
DARK GREENISH BLUE	1	teal (1)
GREYISH YELLOW	1	grey-yellow (1)
YELLOWISH GREEN	1	yellow-green (1)
YELLOWISH	1	golden (1)
VERY-LIGHT YELLOW	1	whitish-yellow (1)
OLIVE BROWN	1	brown-olive (1)
DARK YELLOW	1	mustard (1)
LIGHT REDDISH BROWN	1	red-cream (1)
PINKISH	1	flesh (1)
BRILLIANT RED	1	scarlet (1)
PINKISH BROWN	1	brown-pink (1)
PINKISH GREY	1	grey-pinkish (1)
BLUISH BROWN	1	Blue-brown (1)
BROWNISH PURPLE	1	brownish-purple (1)
LIGHT YELLOWISH RED	1	beige-red (1)
SALT-AND-PEPPER	1	peppered (1)
LIGHT PINK	1	pink-white (1)
LIGHT ORANGE	1	orange-white (1)
LIGHT GREYISH YELLOW	1	grey-beige (1)
LIGHT REDDISH	1	light-reddish (1)
LIGHT BROWNISH	1	light-brownish (1)
ORANGE	1	orange-yellow (1)
PURPLISH BROWN	1	purplish-brown (1)

Table 4. Structure and consistency-related synonyms that were changed to common terms, with frequency of occurrence in parentheses.

STRATIFIED	28496	layers (17473), layer (3972), stringers (1147), streaks (1095), lenses (757), layered (733), mixed (540), trace (532), seams (324), mixture (260), pockets (238), mix (192), stringer (153), interbedded (151), seam (137), zones (113), streak (79), bed (67), lens (64), laminated (61), strips (45), interlayered (36), beds (36), laminations (28), alternating (27), veins (25), varved (23), inclusions (21), strip (19), bands (17), alternate (16), pocket (14), bedded (11), sections (10), interbeds (9), partings (7), layering (7), streaked (6), intermittent (6), intervals (5), vein (5), strata (5), horizons (5), intermixed (4), stripes (4), pods (3), band (3), banding (2), horizon (2), interval (2), lensing (1), intermingled (1), veinlets (1), inclusion (1), intermittently (1)
HARD	2111	solid (1273), dense (398), consolidated (191), harder (130), cement (47), compact (38), compacted (21), denser (6), cementing (2), cementation (1), hard-medium (1)
VERY SOFT	1059	sticky (974), soupy (36), tacky (22), mucky (13), gooey (9), gummy (4), soup-like(1)
SOFT	464	softer (230), greasy (203), clay-like (21), putty-like (7), mushy (2), semiplastic (1)
PLASTIC	347	slick (152), plasticity (146), slippery (49)
FIRM	338	firmer (224), waxy (72), rubber (22), cohesive (4), semistiff (4), semisoft (4), semisolid (3), pliable (2), rubbery (1), semi-firm (1), firmness (1)
WEATHERED	321	rotten (320), weathering (1)
VERY HARD	117	brittle (117)
UNOXIDIZED	85	nonoxidized (85)
FRIABLE	83	crumbley (74), crumbled (3), shattery (2), crumbles (2), crumbling (1), semiconsolidated (1)
OXIDIZED	19	oxidization (10), oxidation (9)
STIFF	17	stiffer (9), semihard (5), medium-hard (3)
UNWEATHERED	15	fresh (15)
MASSIVE	15	homogeneous (15)
LOOSE	4	uncemented (4)

Table 5. Hydrogeology-related synonyms that were changed to common terms, with frequency of occurrence in parentheses.

SATURATED	9559	water (7784), wet (772), water-bearing (415), moist (349), damp (93), flowing (53), frozen (31), watertable (31), ice (9), aquifer (9), moisture (2), frost (2), flows (2), artesian (2), watery (2), flowed (1), waterlogged (1), permafrost (1)
FRACTURED/PERMEABLE	7740	broken (5971), fractures (482), fracture (392), shattered (233), circulation (156), crushed (130), seepage (87), stream (77), porous (56), cracks (37), crevices (15), cracked (14), flakey (11), crack (11), fissures (11), fragmental (9), vuggy (7), permeable (6), brecciated (4), fissile (4), spongy (3), fracturing (2), fissured (2), blocky (2), fragmented (2), crevice (2), porosity (2), pores (2), microfractures (2), pervious (1), honeycomb (1), honeycombed (1), spally (1), platy (1), splintery (1), fissure (1), fault (1)
UNSATURATED	1361	dry (1361)
SALINE	945	salty (945)
EFFERVESCENT/GASSY	108	gas (40), smell (40), sulfur (10), gaseous (6), smelling (4), putrid (2), egg (2), bubbly (1), carbonated (1), methane (1), sulfurous (1)
TIGHT/IMPERMEABLE	82	tight (74), impervious (4), bonded (3), fractureless (1)

Table 6. Words that were retained, with frequency of occurrence.

CLAY	124039
SAND	87191
TILL	85317
GREY	80416
BROWN	59053
GRAVEL	45573
BLUE	29241
FINE	25071
SILTY	18439
YELLOW	18049
SANDY	15483
SOFT	15445
HARD	13175
COARSE	11920
MEDIUM	10689
WHITE	10222
LIGHT	9055
SILT	8515
RED	7992
DARK	7846
TO	7708
FRACTURED	6300
BLACK	5849
CLAYEY	5344
FIRM	4115
SANDSTONE	4000
GRAVELLY	3672
RUBBLE	2694
OXIDIZED	2456
CLEAN	2271
GREYISH	2074
GREEN	2073
DIRTY	1975
FILL	1939
SOIL	1694
SHALEY	1640
COAL	1355
WELL	1259
CARBONATE	1251
LOOSE	1223
BROWNISH	1124
REDDISH	938
CALCAREOUS	645
WEATHERED	592
OLIVE	538
YELLOWISH	511
STIFF	491
ORGANIC	481
CEMENTED	432
ORANGE	420
MOTTLED	396
GREENISH	377
PINK	358
BOULDERY	287
PLASTIC	287
BLUISH	275
UNOXIDIZED	229
NONCALCAREOUS	221
CARBONACEOUS	195
ROUNDED	185
PURPLE	157
PALE	134
ARGILLACEOUS	126
ANGULAR	108
PINKISH	86
MICACEOUS	64
STRATIFIED	37
DOLOMITIC	36
SUBROUNDED	30
GYPSIFEROUS	26
BENTONITIC	24
CHALKY	24
PURPLISH	24
SUBANGULAR	22
FOSSILIFEROUS	21
SALINE	20
MASSIVE	18
KAOLINITIC	17
FRIABLE	14
ARENACEOUS	10
CHERTY	9
PREVIOUS	8
UNWEATHERED	5
OOLITIC	3
VOID	3
NODULAR	3
GLAUCONITIC	3
FELDSPATHIC	3
COALY	2
ARKOSIC	2
CONGLOMERATIC	1
PETROLIFEROUS	1

Table 7. Stratigraphic terminology occurring in the database, with frequency of occurrence.

ODANAH MEMBER	917	RESTON FORMATION	3
MILLWOOD MEMBER	481	MORDEN SHALE	2
SWAN RIVER FORMATION	108	COULTER MEMBER	2
BOISSEVAIN FORMATION	51	GUNTON MEMBER	2
PIERRE SHALE	35	WILLIAMS MEMBER	2
TURTLE MOUNTAIN FORMATION	19	WINNIPEGOSIS FORMATION	2
ASHVILLE FORMATION	9	WINNIPEGOSIS FORMATION	2
RED RIVER FORMATION	6	UPPER FORT GARRY MEMBER	1
STONY MOUNTAIN FORMATION	5	SELKIRK MEMBER	1
WINNIPEG FORMATION	5	STONEWALL FORMATION	1
FAVEL FORMATION	5	BLAIRMORE FORMATION	1
DAWSON BAY FORMATION	4	UPPER RED RIVER FORMATION	1
VERMILLION RIVER FORMATION	4	ASHERN FORMATION	1
RIDING MOUNTAIN FORMATION	3	AMARANTH FORMATION	1
		INTERLAKE GROUP	1

have been applied to a single entity. Many classification schemes have been published, but the approach taken here was to build a classification that could capture as much information as possible from the database as it stands. To some extent, additional terminology was added to accommodate potential future usage, but this was limited in order to avoid an unnecessarily complicated system with terms that might never be used. It should be noted however, that the terminology could readily be extended to accommodate geology in other areas, or increasing geological sophistication in the drilling community.

All information herein makes reference to a stratum intersected by a drill hole. The entities in the database therefore are identified on the basis of a site identifier, a stratum sequence number starting at 1 at the top, a depth to top of the stratum, and a depth to the stratum's lower contact (table 8). A distinction is then made whether the interval has been described or not, whether it is a previous well, or whether it is a geological material, either sediment or rock. If it is a sediment or rock that was described, a basic description was parsed into five variables (table 9). If more detailed information is provided, typically by a geologist, this more detailed information was placed into supplementary tables that provide an extended description of the geology (tables 7 and 10 to 15).

Rules of precedence were required, and typically it was only the first piece of information on a topic in a stratum description that was parsed under the relevant attribute. The other information after the first word can now only be retrieved by searching the original descriptions or, preferably, the spell-checked original description. An exception to this rule is some cases of composite adjectives, where it is the final adjective that may take precedence. There also are cases where a composite noun is modified by a composite adjective, and in this case there is a tendency for precedence to be given to the first noun, but sometimes the first, but elsewhere the second, adjective. Also, judgment was used to identify the most noteworthy of multiple observations with respect to geology.

Some compound nouns were parsed as a noun and a modifier, where geological knowledge implies that the two entities were blended in a homogeneous manner. Other compound nouns imply to a geologist a heterogeneity that would have resulted in the two nouns being parsed to the 'lithology' and the 'interbedded with' fields. For example, 'silt and clay' implies homogeneity to a geologist, where 'limestone and shale' implies heterogeneity.

MATERIAL

The most general level of stratum classification is the material. From driller's records, we classify the materials as soil/sediment, rock, ice, water, unclassified, or as a preexisting well (table 9). The term soil/sediment accommodates the engineering view that anything above bedrock is soil, as well as the geologists' term 'sediment', including deposits described as unconsolidated, unlithified, drift, sediment, surficial deposits, or overburden. 'Rock' encompasses all hard rock as well as in situ weathered rock, excluding boulders. This includes deposits referred to as 'bedrock', 'caprock', and 'stone'. 'Ice' and 'water' were added to the table of options in anticipation of data from lakes, frozen lakes, or glaciers. Unclassified intervals are those for which no information was available, such as blanks where no sample was recovered from the borehole, or where the driller recorded 'pit', 'open pit', 'no log', or an adjective with no noun. Previous wells that were being deepened or re-drilled were described using terminology such as 'old well', 'well pit', 'existing well', 'dug well', 'previously drilled', 'cement crib', 'cribbing', or 'casing'.

Table 8. Designation of the stratum.

Site ID	Unit #	From depth	To depth	Material
				unclassified
				previous-well
				ice
				water
				soil/sediment
				rock

Table 9. Basic rock or sediment description.

Principal colour	Textural modifier	Texture	Lithology	Interbedded with
black	clayey	clay	fill	fill
blue	silty	silt	topsoil	topsoil
brown	sandy	sand	peat/organic	peat/organic
green	gravelly	very-fine	silt/clay/fines/mud	silt/clay/fines/mud
grey	bouldery	fine	sand	sand
multicoloured	very-fine	medium	gravel	gravel
olive	fine	coarse	boulder(s)	boulder(s)
orange	medium	very-coarse	till/diamict	till/diamict
pink	coarse	small	soil/sediment	soil/sediment
purple	very-coarse	medium	coal	coal
red		large	evaporites	evaporites
white		very-large	carbonate	carbonate
yellow			shale/mudstone	shale/mudstone
			sandstone	sandstone
			conglomerate	conglomerate
			igneous/metamorphic	igneous/ metamorphic
			rock	rock
			void	void

LITHOLOGY

Following designation of the general material type, as much additional information as possible was captured for every stratum designated soil/sediment or rock. The first two distinct lithological entities described for each stratum were assigned to the 'lithology' and 'interbedded with' attributes (table 9).

Sediments referred to as fill included such descriptions as 'backfill', 'roadbed', 'concrete', 'dyke fill', 'pavement', and 'driveway'. These are human-built deposits of limited thickness at land surface.

Deposits referred to by the driller as topsoil include references to 'soil' at surface, and typically are the thin, organic-rich A-horizon deposits, including 'topsoil', 'loam', 'sod', 'dirt', and 'black earth'. These occurrences are distinct from the engineering usage of the term 'soil', where soil refers to everything above bedrock. The latter we call undifferentiated soil/sediment. Organic sediments we refer to as peat/organic include those described as 'peat', 'muskeg', 'swamp decay', 'organic', 'quag', and 'swamp'.

Driller's descriptions of clay or silt were lumped together as fines, also known as mud in geological terminology. This was done in recognition of the likelihood that silt deposits, especially fine silt, would normally be described as clay, even by trained geologists. Designation as clay or silt, however, is retained in the texture variable (also in table 9). Sediments referred to as sand include those described as 'sand and gravel', on the basis of our opinion that sand and gravel deposits consist mostly of sand, and therefore were classified as gravelly sand. Sediments that we refer to as gravel occur in many original descriptions as various synonyms, as already discussed. The term 'boulder(s)' was retained, although it could be argued that these intervals could be labelled 'rock', while still using the word 'bouldery' in an adjective field. The term 'rubble' is commonly used, but for the sake of simplicity, was changed to 'angular boulders'.

The lithological theme that required the most judgment was selection of driller's descriptions that are regarded by the authors as likely to be glacial diamicts, in most cases deposited as till by glacial ice. Diamicts occurring as diamictites or diamictons are relatively unsorted sediments containing a range of grain sizes from clay to gravel. In the database, the term till/diamict was used to broaden recognition of the term, although other terms exist in the formal scientific and technical literature, such as matrix-supported gravel, morainal material, and loam. Strata coded as till/diamict included those described as 'clay and stones', 'clay and boulders', 'clay and rocks', 'gravel and clay', 'hardpan', and 'boulder clay'. In other words, any stratum that included reference to both clay and gravel was designated as till. Although many till deposits in the region are very silty, descriptions of gravelly silt were not changed to till on the basis of the expectation that a driller who uses the word silt likely possesses, at least in the majority of cases, a relatively high level of geological sophistication and therefore is more likely to be distinguishing an actual silt deposit, perhaps with thin gravelly interbeds, from a till. However, even experienced geologists frequently have difficulty distinguishing till from poorly sorted silty deposits in drillhole samples, and doing so from washed cuttings and drilling rate presumably is also difficult.

Actual interpretation of the driller's description, and hence conversion to something that was not said by the driller was strictly avoided. It could even be argued that no interpretation was made, and reformatting was limited to synonym substitution, if it is accepted that 'clay with stones' is a synonym of 'till'. For example, clay interbedded with limestone was left as such, rather than intervening to designate the limestone a boulder, and translating the description to something like carbonate-rich bouldery till. This was done in recognition that a deposit described as clay interbedded with limestone could be soft shale interbedded with limestone, and designation as till would have been significantly in error. Furthermore, sand was coded as sand, even when interbedded with sandstone and/or shale, and clay was left as clay, even if interbedded with sandstone and/or shale.

Rocks that we coded as coal were described by the driller as either 'coal' or 'lignite'. In a manner similar to the line of reasoning discussed above with respect to silt and clay, it was our opinion that anhydrite deposits, if present, would likely have been referred to as gypsum, so the term 'evaporites' was used for deposits described as 'gypsum', 'anhydrite', or 'gyprock'. Similarly, for carbonates, it was assumed that many strata described as limestones are in fact dolomites, so the term 'carbonate' was used, although the driller's specification of limestone or dolostone was retained as carbonate modifier.

Shales, mudstones, siltstones, and claystones were lumped under the term 'shale/mudstone'. These included those described as 'shale', 'bentonite', 'soapstone', and 'slate', although the latter two terms were included among unmetamorphosed rocks only after inspection of the context. Rocks referred to as sandstone are in nearly every case described as such in the original data, but also include 'silica sand' where included in a rock sequence. The term 'conglomerate' is rarely reported in the region, but was retained.

Also occurring in rock sequences are voids, which are reported as 'cavity', 'sinkhole', 'cavern', 'void', 'cave', 'open cavity', 'free fall', 'empty space', or 'large crevice'. This is not unexpected in the karst terrain found in the region.

All igneous and metamorphic rocks were lumped in one category at the level of basic description, including rocks described as 'granite', 'igneous bedrock', or 'Precambrian rock'. Distinction of intrusive and high-grade-metamorphic rocks, such as granite, gneiss and schist, from low-grade-metasedimentary and metavolcanic, such as greenstones, was retained in another field.

Where two distinct lithologies were indicated by words such as ‘and’ or ‘layers’, the second named lithology was placed in the ‘interbedded with’ variable, although no provision was made for description of the interbedded material. A description such as ‘till and sand’, for example, was parsed as till interbedded with sand, and the till was given no textural modifier related to the interbedded sediment.

TEXTURE

Adjectives that provide information about grain size were divided into three variables. The principal modifier, here referred to as ‘Texture’ (table 9) includes three groups of modifiers. The words ‘clay’, ‘silt’, and ‘sand’ are intended to be modifiers of fines and till/diamict, as in ‘silt diamict’. The terms ‘very fine’, ‘fine’, ‘medium’, ‘coarse’, and ‘very coarse’ are primarily intended as modifiers for sand, but also can be used for silt and gravel. The terms ‘small’, ‘medium’, ‘large’, and ‘very large’ also were included in this variable for possible future use with respect to boulders.

An agreed-upon set of definitions for these terms would ideally be adopted for the future, but it is recognized that the existing database likely was built using several slightly differing schemes. For example, the exact definition of clay and sand, with respect to quantitative grain size cut-offs, varies slightly among the many classification schemes currently in use.

Also appearing in the basic lithological description (table 9) is a second texture-related adjective, here called the textural modifier. The words ‘clayey’, ‘silty’, ‘sandy’, ‘gravelly’ and ‘bouldery’ are intended to be used as modifiers with or without use of another word under ‘Texture’. For example, ‘silty sand’ is accommodated, although future usage should encourage the dominant sand texture class to be identified, as in ‘silty very fine sand’. Usage such as ‘silty sand diamict’ also is accommodated. The words ranging from ‘very fine’ to ‘very coarse’ also are included here to permit usage such as ‘coarse silt’.

Where boulders were mentioned in a stratum description, this observation took precedence as a textural modifier. In the case of till initially described as some combination of clay and gravel, if the gravel fraction was described as something like an odd pebble, a few pebbles, or a trace of gravel, the bed was classified a clay till/diamict. Deposits described as ‘clay, gravelly’ were simply classed as till with no textural modifier, because any till/diamict deposit has a gravel fraction. In contrast, ‘till, gravelly’ was parsed with a textural modifier, due to the implication that the deposit contains more gravel than the amount adequate for designation as till. Similarly, ‘till, clayey’, ‘till, silty’, or ‘till, sandy’, were parsed with a textural modifier. Multiple textural modifiers were disregarded due to ambiguity, although the information remains available in the original text.

In addition to the above, there is additional textural information in the database related to the range of sand texture, as well as sorting and clast shape. In order to minimize the complexity of the basic description, and because this information typically is only available for strata described by a geologist, the information was placed in a subordinate table referred to as an extended sediment description (table 10).

The terms ‘very fine-to’, ‘fine-to’, ‘medium-to’, ‘coarse-to’ are meant to be used in combination with

Table 10. Extended sediment description.

Weathering	Structure	Consistency	Lithological modifier	Carbonate modifier	Textural range	Sorting	Roundness 1	Roundness 2
oxidized	massive	loose	carb/evap&clstcsedrx-rich	noncalcareous	very-fine-to	well-sorted/ graded	very-angular-to	very-angular
unoxidized	stratified	friable	carb/evap&ig/met-rich	slightly-calcareous	fine-to-	poorly-sorted/ graded	angular-to	angular
		very soft	carb/evap-rich	calcareous	medium-to-		subangular-to	subangular
		soft	clstcsedrx&ig/met rich	strongly-calcareous	coarse-to-		subrounded-to	subrounded
		plastic	clstcsedrx-rich				rounded-to	rounded
		firm	ig/met-rich					well-rounded
		medium stiff	mixed-lithology					
		stiff	very-carb/evap-rich					
		very stiff	very-clstcsedrx-rich					
		hard	very-ig/met-rich					
		very hard	clean					
			dirty					
			authigenic-xl-bearing					
			organic					
			wood-bearing					

'fine', 'medium', 'coarse', and 'very coarse', in order to specify a textural range. This primarily is intended for sand, but also can be used for silt and gravel. Sorting and roundness/angularity rarely are described in the database. Where the information is provided, sorting was categorized as 'well sorted/graded' or 'poorly sorted/graded' (table 10). A range was used for roundness, with the first variable giving the more angular limit of the range, including 'very angular-to', 'angular-to', 'subangular-to', 'subrounded-to', and 'rounded-to', and the second roundness variability giving the more rounded limit, either 'very angular', 'angular', 'subangular', 'subrounded', 'rounded, or 'well rounded'. A full set of terms in the second roundness variable permits one-term descriptions, such as 'very angular gravel'.

SEDIMENT STRUCTURE/COMPOSITION

In many extended descriptions of sediments there is additional information on structure and composition (Table 10). The structure of sediments was classified as either massive or stratified within a defined unit. The term massive is rarely specified, but a large number of words convey the presence of stratification. The consistency of sediment, where described, was classified as either 'loose', 'friable', 'very soft', 'soft', 'plastic', 'firm', 'medium stiff', 'stiff', 'very stiff', 'hard', or 'very hard'.Information on sediment composition includes indications of weathering, whether 'oxidized' or 'unoxidized', and carbonate content, classified as either 'noncalcareous', 'slightly calcareous', 'calcareous', or 'strongly calcareous'.

The lithology of the gravel fraction in sand and gravel or till deposits commonly is reported in well descriptions. A scheme was developed to classify these gravels based on varying proportions of composition: 1) carbonate and evaporite rocks; 2) clastic sedimentary rocks including shale, sandstone, and conglomerate; and 3) igneous and metamorphic rocks. A composition classification is contained in table 10 under the "Lithological modifier". Where one composition class strongly dominates (>~90%), the composition is prefaced by 'very', as in 'very carbonate-rich'. Where there is an indication that one class is dominant (>~50%), that class is named, as in 'carbonate-rich. Where two lithologies are named, implying that the two classes make up more than ~75% but neither of them exceed 50%, a designation such as 'carbonate-and-shale-rich' (which is listed as 'carbonate/evaporite' and 'clastic-sedimentary-rock-rich' in the generic language of Table 10) was used. In cases where each of the three classes exceeds a quarter, but none exceed a half, a term such as mixed-lithology would be used. Other indications of sediment composition included clean, dirty, authigenic-crystal-bearing, organic, and wood-bearing.

EXTENDED ROCK DESCRIPTION

Numerous sedimentary rock descriptions include information more detailed than that suitable for the basic description, including information on weathering, detailed lithological attributes, and carbonate content. This information is rarely found. Where information on weathering of sedimentary rocks is found in drill records, it was parsed as either weathered or unweathered (table 11).

Detailed information on lithology that could readily be parsed as a lithological modifier (table 11) included the terms 'arenaceous', 'argillaceous', 'arkosic', 'bentonitic', 'carbonaceous/bituminous', 'chalky', 'cherty', 'concretionary/nodular', 'conglomeratic', 'dolomitic', 'feldspathic', 'ferruginous', 'fossiliferous', 'glauconitic', 'gypsiferous', 'kaolinitic', 'micaceous', 'oolitic', 'petroliferous', 'shaley', 'silica/siliceous', and 'sulphide-bearing'.

In a manner similar to that used for sediments, provision was made to permit description of sedimentary rocks as either 'noncalcareous', 'slightly calcareous', 'calcareous', or 'strongly calcareous' (see "Carbonate modi-

Table 11. Extended sedimentary rock description.

Weathering	Lithological modifier	Carbonate modifier
weathered	arenaceous	noncalcareous
unweathered	argillaceous	slightly-calcareous
	arkosic	calcareous
	bentonitic	strongly-calcareous
	carbonaceous/ bituminous	limestone
	chalky	dolostone
	cherty	
	concretionary/ nodular	
	conglomeratic	
	dolomitic	
	feldspathic	
	ferruginous	
	fossiliferous	
	glauconitic	
	gypsiferous	
	kaolinitic	
	micaceous	
	oolitic	
	petroliferous	
	shaley	
	silica/siliceous	
	sulphide-bearing	

fier" in table 11). In addition, the words 'limestone' and 'dolostone' are included under this attribute as modifiers of carbonate.

Among igneous and metamorphic rocks, indications of weathering were frequently reported, and these were captured as either 'weathered' or 'unweathered'. Where possible, these rocks were further subdivided as either intrusive/high-grade metamorphic or low-grade metasedimentary/metavolcanic (table 12).

Table 12. Extended igneous/metamorphic rock description.

Weathering	Lithological modifier
weathered	intrusive/high-grade-metamorphic
unweathered	low-grade-metasedimentary/metavolcanic

COLOUR

All colour terminology interpreted from the well descriptions was coded according to the Munsell Book of Colour, a system that can readily be converted to other systems. The most important colour information is the single word such as 'brown' or 'grey' that was placed in an attribute called 'principal colour' (table 9). To capture other colour information, a colour such as 'light yellowish brown' was parsed into three variables (principal colour, principal value/chroma, and principal colour modifier; tables 9 and 13). In addition, in the rare cases when it was described, the pattern of multiple colours, whether 'mottled' or 'speckled', was coded. If the first two colours were distinct, the word 'and' was used as a colour link, whereas gradational colour relationships were indicated using the word 'to'. This allowed a description such as 'mottled light yellowish brown to dark greyish green' to be fully captured. The term 'salt and pepper' was recorded as 'speckled'.

Multiple colours were handled arbitrarily by assigning the first named colour to principal colour and the second named colour to the secondary colour. It is acknowledged that there commonly will be ambiguity in such cases, although these concerns relate only to a small number of strata, where multiple lithologies and multiple colours are mentioned. Table 14 presents examples of hypothetical cases that range from clear to ambiguous when parsed. While the first example in table 14 is clear, the second will be ambiguous when parsed, if not misleading, due to the tendency of the user to assume that a single colour relates to the first listed lithology. The third example, 'red shale and blue clay' will be ambiguous, but also misleading with respect to the tendency for the user to read the parsed data as red and blue shale interbedded with clay. Other examples in the table indicate, for example, capture of only the first two reported colours. The level of ambiguity in these situations is considered acceptable due to their low frequency of occurrence, and also due to recognition of the database complexity that would have been required to avoid ambiguity.

HYDROGEOLOGY

Information related to hydrogeology was captured in three variables (table 15). Strata were classified where possible as either 'unsaturated' or 'saturated' with water. Indications of gases based on effervescence or odour were captured as 'effervescent/gassy' under a solutes attribute, as were indications of salinity. Indications of permeability were classified as either 'fracture/permeable' or 'tight/impermeable'.

STRATIGRAPHY

Stratigraphic nomenclature was captured (table 7), except where it made reference to the provenance of a gravel, such as 'gravel composed of Pierre shale'.

QUALITY ASSURANCE

A key issue is the level of reliability that can be granted to the database. In answering this concern, a crucial distinction has to be made between applications that make reference to one site, versus applications that make reference to many sites in a manner that takes the location into consideration. When the information from one site is being inspected, there is little to be gained by referring to the parsed information rather then the original description, and a high level of caution is required in case the location is in error or the geological description is misguided. In contrast, however, are cases where the data from a large number of sites can be viewed, with location taken into consideration. By enabling the presentation of several hundred sites in map or section view, with the aid of x, y, and z coordinates along with categorized geological information that can be depicted using colours or symbols, a powerful means to assess reliability is obtained, because anomalous sites will differ from adjacent sites in a manner that is unreasonable from a geological perspective. Applications of the database are now indicating a surprisingly low frequency of apparent location errors or mistaken geological categorizations.

RECOMMENDATIONS

1. It is recommended that the parsed information be audited by members of the group responsible for maintenance of GWDrill, so that they may make their own conclusions regarding its validity.
2. Ongoing changes should be made to the classifica-

Table 13. Extended colour description.

Colour pattern	Principal value/chroma	Principal colour modifier	Principal colour (*from main table*)	Colour link	Secondary value/chroma	Secondary colour modifier	Secondary colour
mottled	brilliant	blackish	black	and	brilliant	blackish	black
speckled	dark	bluish	blue	to	dark	bluish	blue
	dusky	brownish	brown		dusky	brownish	brown
	light	greenish	green		light	greenish	green
	medium	greyish	grey		medium	greyish	grey
	moderate	olive	multicoloured		moderate	olive	multicoloured
	pale	orange	olive		pale	orange	olive
	strong	pinkish	orange		strong	pinkish	orange
	very-dark	purplish	pink		very-dark	purplish	pink
	very-dusky	reddish	purple		very-dusky	reddish	purple
	very-light	yellowish	red		very-light	yellowish	red
	very-pale		white		very-pale		white
	weak		yellow		weak		yellow

Table 14. Examples: parsing of colour information.

Hypothetical original description	Principal colour	Colour link	Secondary colour	Lithology	Interbedded with
red shale	red			shale	
shale and blue clay	blue			shale	clay
red shale and blue clay	red	and	blue	shale	clay
red to blue clay	red	to	blue	clay	
red and blue clay	red	and	blue	clay	
shale and red to blue clay	red	to	blue	shale	clay
shale and red and blue clay	red	and	blue	shale	clay
red to blue clay and green to black shale	red	to	blue	clay	shale
red to blue clay and green and black shale	red	to	blue	clay	shale
red and blue clay and green to black shale	red	and	blue	clay	shale
red and blue clay and green and black shale	red	and	blue	clay	shale

Table 15. Extended hydrogeological description.

Water	Solutes	Permeability
unsaturated	effervescent/gassy	fractured/permeable
saturated	saline	tight/impermeable

tion and its implementation on the basis of audits, test applications, and user consultations. These changes could initially be made by the authors during a transition period, but this responsibility should be transferred to the managers of the database at the earliest opportunity.

3. Following satisfactory audit and user consultations, the revised parsed information should be attached to the existing database using site ID and stratum sequence number as unique identifiers for the strata. At least the basic description, consisting of five variables, should be attached, if not the more detailed tables as well.

4. Data acquired in the future should be parsed to the adopted structure. Various means to accomplish this task on an ongoing basis should be tested. The simplest means of doing so, in our opinion, would be for new stratum descriptions, following correction of spelling, to be compared to the existing database, and for the existing parsed information to be duplicated where a match to a previously acquired record is found. In cases of newly submitted stratum descriptions that have never appeared before, the record could be referred to a geologist for interpretation, with the aid of a checklist of attributes and terminology. Although software that would process all conceivable wording could potentially be developed, this is not seen as a favourable option from the point of view of cost/benefit analysis.

5. The potential should be considered for water well drillers to utilize: A) a GPS receiver capable of obtaining adequate location data, and B) a checklist of attributes and words with which to describe strata. Doing so would require training and interaction with the drilling community on an ongoing basis, and would require provision for unrestricted comments to accompany the coded information.

ACKNOWLEDGMENTS

This work was done in increments over several years, and interaction over that time with R. Bezys, R. Betcher, L. Donnelly, R. Eilers, R. Fulton, L. Gray, C. Kaszycki, G. Keller, G. Matile, L. Murray, and F. Render is gratefully acknowledged.

Geoscience Terminology Development for the National Geologic Map Database

By Stephen M. Richard[1], Jonathan Matti,[2] David R. Soller[3]

[1]Arizona Geological Survey
416 W. Congress # 100
Tucson, AZ 85701
Telephone: (520) 770-3500
Fax: (520) 770-3305
email: Steve.Richard@azgs.az.gov
[2]U.S. Geological Survey
e-mail: jmatti@usgs.gov

[3]U. S. Geological Survey
908 National Center
Reston, VA 20192
e-mail: drsoller@usgs.gov

ABSTRACT

Science language vocabularies are being constructed for use in the U.S. National Geologic Map Database (NGMDB), to provide geologic data to users in government, business, academia, and the general public in a consistent data structure, using consistent descriptive terminology or science language. These terms and definitions are used to classify observable or inferred facts or events, and to assign values for properties in descriptions. The vocabulary makes pre-defined terms available to apply in descriptions without having to reconstruct the entire description denoted by the vocabulary term. Terms within a controlled vocabulary may be *atomic* or *analyzable*. The atomic terms are irreducible in the database environment; they are either numbers, or terms defined by narrative text. The text definition is meant for a human user, and is not amenable to automated analysis. Analyzable terms are specified by combinations of properties and relationships according to some formal description structure. Analyzable terms may be compared, using the values of the properties in their formal description. The atomic terms in a vocabulary may represent complex concepts that could be represented by description schema. Implementation of formal descriptions for terminology associated with a concept allows users to develop alternate or more specific terminology for concepts that may interoperate with other systems by using automated description comparisons (Description Logic).

The unifying feature of the NGMDB is a common conceptual model and terminology system. To function as a part of the NGMDB, individual databases must conform to, or interoperate with, the common conceptual model and shared terminology. The simplest route to NGMDB conformance is to use the standard science vocabularies. If desired, users can build their own vocabularies, but in order to be NGMDB conformable, user-defined vocabularies will need to: 1) define the scope of their vocabulary; 2) provide text definitions of each vocabulary term; 3) build formal definitions rooted in the standard science vocabulary, using description schema from the common conceptual model, and a standard, well documented syntax; and 4) provide a correlation of each term with the most specific subsuming term from the standard vocabulary.

INTRODUCTION

Science language vocabularies are being constructed for use in the National Geologic Map Database (NGMDB), a distributed, federated information system that is planned for integrating geologic data provided by state and federal geological surveys. The goal is to provide geologic data to users in government, business and academia in a consistent data structure, using consistent descriptive terminology. A conceptual model referred to as NADM-C1, proposed as a standard data structure for geologic map data, has been developed by the Data Model Design Team of the North American Data Model Steering Committee (NADMSC) and is under review (Data Model Design Team, 2003). The National Geologic Map Database Project will use this model as a foundation for database development. A Science Language Technical Team of the NADMSC has developed a draft science language that is currently in review as well (http://geology.usgs.gov/dm/steering/teams/language/charter.shtml), and the NGMDB will use this vocabulary with extensions as necessary. The use of a clearly defined vocabulary, with definitions developed to avoid ambiguity in the use of terms, is essential for consistent communication of science concepts to a wide variety of data users with varying geologic expertise.

Implementation of any geoscience information system requires a collection of science language terms and definitions to populate descriptions of rocks, geologic units, and structures (Giles and others, 1997). A collection of terms, their definitions, and relationships between the terms (especially hierarchical relationships) are referred to here as a controlled vocabulary or terminology system (de Keizer and Abu-Hanna, 2000, de Keizer and others, 2000). The data model implemented by a database dictates the kinds of controlled vocabularies required to actually populate the database. For example, the NADM-C1 models 'WeatheringCharacter' as a property of a geologic unit. This property is specified by a term from a science language vocabulary, thus requiring a controlled vocabulary of terms that specify different weathering character values. Controlled vocabularies are used to populate fields in a database in order to make possible the clear and unambiguous communication of database content. The fundamental difficulty is the variability in usage of many common geoscience terms. Consider the various usages of the term 'shale':

- "*Shale* is a laminated or fissile claystone or siltstone....To those claystones which are neither fissile nor laminated but are blocky or massive, the term mudstone may be applied." (Pettijohn, 1957, p. 341)
- "Shale...is rock that splits parallel to bedding into thin flakes or plates." (Williams and others, 1954, p. 316);
- "In this book, the author uses the term shale for all sedimentary rocks composed dominantly of mud-size (<0.6 mm) particles." (Boggs, 2001, p. 153).
- "A widely used and often loosely used field term for mudstone which shows a conspicuous fissility on weathering. It is somewhat unsatisfactory in that the weathering state must play a part in its recognition and it cannot be consistently used in comparing rock at outcrop with, say, that of a borehole core" (Collinson and Thompson, 1989, p. 54).
- According to Potter and others (1980, table 1.2), "mudshale" and "clayshale" are terms restricted to rocks that are laminated (<10 mm) and that have specific ranges of clay-size constituents: 33-65% (mudshale) and 66-100% (clayshale);
- "*Shale* is the term used for fissile mudrock and, more generally, for the entire class of fine grained sedimentary rocks that contain substantial quantities of clay minerals." (Blatt and others, 1980, p. 374). These authors emphasize fissility as a core requirement for the definition of shale (also see Folk, 1968).

It should be apparent that the use of shale as a rock name in a database, without some connection with the definition of shale the user is implying, does not communicate unambiguously what kind of rock is being described. Standardized terminology is being developed in order to avoid these sorts of problems.

In this paper, the NADM conceptual data model (NADM-C1) is first reviewed because it defines the framework within which the NGMDB science language is being developed. Then the design of the NGMDB science language is described.

NADM CONCEPTUAL MODEL

The NADM-C1 for geologic map information, version 1.0 (Data Model Design Team, 2003) is designed to be a technology-neutral conceptual model that can form the basis for a web-based interchange format using evolving information technology (such as Extensible Markup Language, or XML), and for the design of database implementations that can interoperate using the interchange format. The intended purpose is to allow the sharing of geologic information independent of logical and physical implementations. NADM-C1 is a model of geoscience concepts and the relationships between them, with special emphasis on concepts related to information presented on geologic maps (Richard, 1999 and in press; Brodaric and Gahegan, 2000; Brodaric and Hastings, 2002; Soller and others, 2002).

The top level of NADM-C1 is 'NADMUniverse,' representing all concepts in the model. Three subclasses of NADMUniverse are represented: GeologicConcept, Metadata, and GeologicRepresentation (fig. 1). The metadata part of NADM-C1 is outside the scope of this paper and will not be discussed. Concept names in the text written using the Courier font refer to NADM-C1 model elements, and concept names in plain text refer to geoscience concepts in general. The term 'concept' as used here represents the notion of any mental phenomena that human beings use in their internal representation of the world. Webster's dictionary (http://www.m-w.com/) uses the terms 'idea' and 'object of thought' to convey the meaning of 'concept.'GeologicConcept includes concepts specific to the domain of geoscience knowledge. Figure 1 shows a generalized version of the modeled GeologicConcept hierarchy. The top-level concept that pertains to geologic maps and geologic information (loosely defined) is called GeologicConcept. The names in the boxes beneath the major concept names represent additional concepts that inherit from GeologicConcept. In other words, GeologicStructure is a GeologicConcept and Contact is a GeologicStructure and, therefore, a GeologicConcept. Table 1 defines major geologic concepts in NADM-C1.

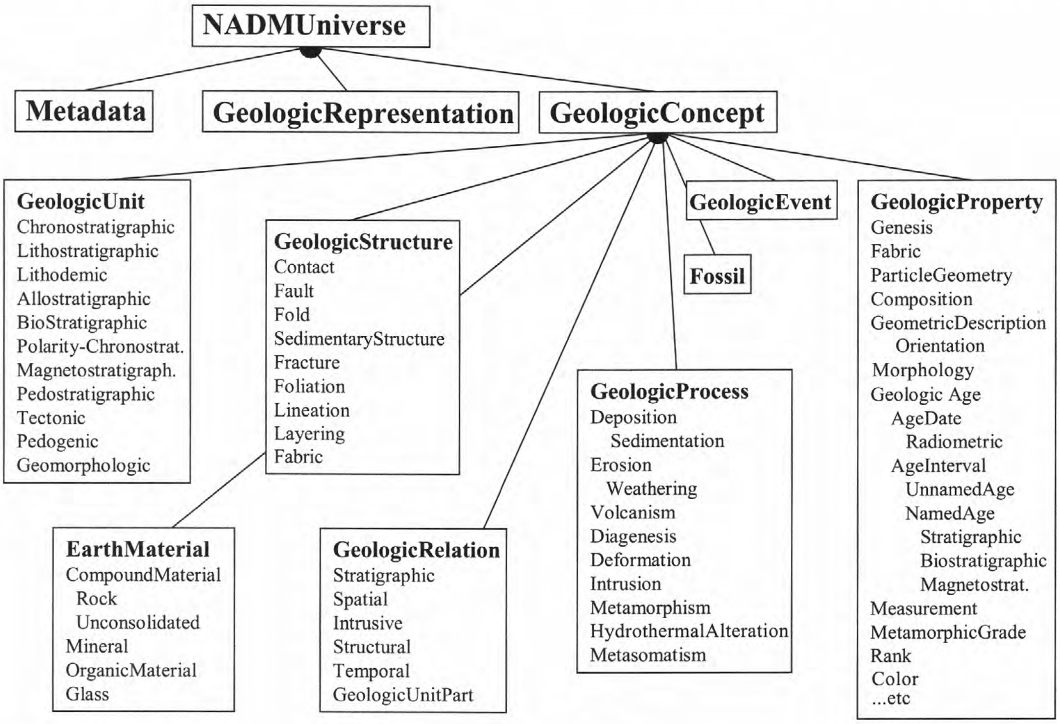


Figure 1. High-level geologic concept hierarchy for NADM-C1.

Table 1. Major sub-concepts of GeologicConcept in NADM-C1.

Concept	Scope and rationale
EarthMaterial	A naturally occurring substance in the earth. EarthMaterial represents substance, and is thus independent of quantity or location. Ideally, an EarthMaterial is defined strictly on the basis of physical properties, but because of standard geologic usage, genetic interpretations commonly enter into the definition as well. Does not include melted rock (magma or lava). Many concepts related to water or petroleum have not been modeled in this version.
Fossil	The remains, trace, or imprint of a life form that has been preserved in an EarthMaterial, and that demonstrates evidence of having been changed from its original biogenic form. Fossil is distinguished from biologic remains or biogenic sedimentary structure based on evidence of having been converted incipiently or substantially into a modified version of the original biogenic form or structure. Although the passage of time is implicit in the definition of Fossil, no constraint is placed on the amount of time necessary to become a fossil.
GeologicEvent	An identifiable event during which one or more geologic processes act to modify a geologic entity (Earth material, geologic unit, or geologic structure). A GeologicEvent may have geologic properties such as a specified GeologicAge and Geologic Environment. An example might be a cratonic uplift event during which erosion, sedimentation, and volcanism occur.
GeologicProcess	A function, possibly complex, that acts on one geologic entity to produce another geologic entity at a later time. Process is time independent; some GeologicProcesses are observable today in the field or in the laboratory, others can only be inferred from observing the results of the process. Processes take one or more of EarthMaterial, GeologicUnit, or GeologicStructure as input and have one or more of EarthMaterial, GeologicUnit or GeologicStructure as output.
GeologicProperty	An inherent feature used to characterize a GeologicConcept. Some examples include physical properties, color, age, grain shape, metamorphic grade, and weathering character.

GeologicRelation	Any of a wide variety of relationships that can exist between two or more GeologicConcepts. For example, the GeologicRelation 'intrudes' is a relationship between an intrusive igneous rock and some host rock. Includes spatial, temporal, sequence, correlation, and parent/child relations. Many of the relationships in NADM-C1.0 (particularly attribute links and parent-child links) are not explicitly modeled as kinds of GeologicRelation, but are represented as associations in the model schema.
GeologicStructure	A configuration of matter in the Earth based on describable inhomogeneity, pattern, or fracture. The identity of a geologic structure is independent of the material that is the substrate for the structure. Properties like 'clast-supported', 'matrix-supported', and 'graded bed' that do not involve orientation are considered kinds of GeologicStructure because they depend on the configuration of parts of a rock body. Includes sedimentary structures.
GeologicUnit	A naturally occurring body of material distinguished from adjoining material on the basis of content (lithologic or fossil), inherent attributes, physical limits, geologic age, or some other property or properties (adapted from North American Commission on Stratigraphic Nomenclature, 1983, p.22; http://www.agiweb.org/nacsn/code2.html). Corresponds to 'stratigraphic unit' in the North American Stratigraphic Code. Commonly used properties include composition, texture, included fossils, magnetic signature, radioactivity, seismic velocity, and age. Sufficient care is required in defining the boundaries of a unit to enable others to distinguish the material body from those adjoining it (North American Commission on Stratigraphic Nomenclature, 1983).

GeologicRepresentation includes concepts related to the representation of geoscience knowledge, and this paper deals particularly with one representation concept—GeologicVocabulary. A geologic vocabulary is a collection of controlled concepts, referred to here as terms. Each of these is associated with a preferred name, and terms are usually organized in some logical fashion such as a hierarchy. The preferred name for a controlled concept in any particular vocabulary provides a standard means of identifying the associated GeologicConcept in that context, and can be thought of as a proxy for the collection of property values and relationships specified in the definition of the controlled concept. Examples of geologic vocabulary include a collection of standard rock types, a stratigraphic lexicon, or a geologic time scale.

A description is a collection of property values and relationships that apply to some thing (in the most general sense), specifying the nature of the thing such that it may be identified. A term from a science language vocabulary implies the properties and relationships specified in its normative description. The normative description may be a simple body of text, in which case the controlled concept is not computationally analyzable. Other normative descriptions may be specified by a description scheme that establishes the properties and relationships that may be assigned values to build a description. Such descriptions may be analyzed using software inference tools to determine subsumption and class membership relationships. For example, a rock is defined as 'a compound material that is consolidated'. The description scheme for compound material specifies that (among other things) a compound material has a 'consolidation degree' property. This property is specified by a term from the consolidation degree terms vocabulary. Given a collection of compound material descriptions, all instances of rock could be identified by selecting for those with a consolidation degree property value that is 'consolidated' or some term that is a child of 'consolidated' (such as 'strongly indurated').

Geologic vocabularies play a central role in the application of NADM-C1 for storing geologic information. Terms from geologic vocabularies are used to classify observable or inferred facts or events (phenomena), and to describe them by assigning values for various properties (fig. 2). Controlled vocabularies of rock names, kinds of geologic structures, or geologic units (stratigraphic lexicon) are used to classify observations to be recorded in a database. The vocabulary makes the definitions of controlled concepts available to apply in other descriptions without having to reconstruct the entire description denoted by the term. For example, in the description of a geologic unit, a constituent may be identified simply as 'granite'—a term from a standard lithology vocabulary, which has an associated normative description that follows the description scheme for a compound material as specified in the NADM-C1 model. Because the controlled term 'granite' is linked to a description, no other data entry is necessary in order to infer that the properties of 'granite' (consolidation, color, composition, etc.) apply to some part of the geologic unit. The NADM-C1 model identifies numerous geologic properties that are used in description or definition of geologic concepts. Each of these properties requires an associated vocabulary of terms to specify possible values for that property in a database.

As an example, consider the following text from a published geologic report:

> "The diabase is typically dark gray or green-

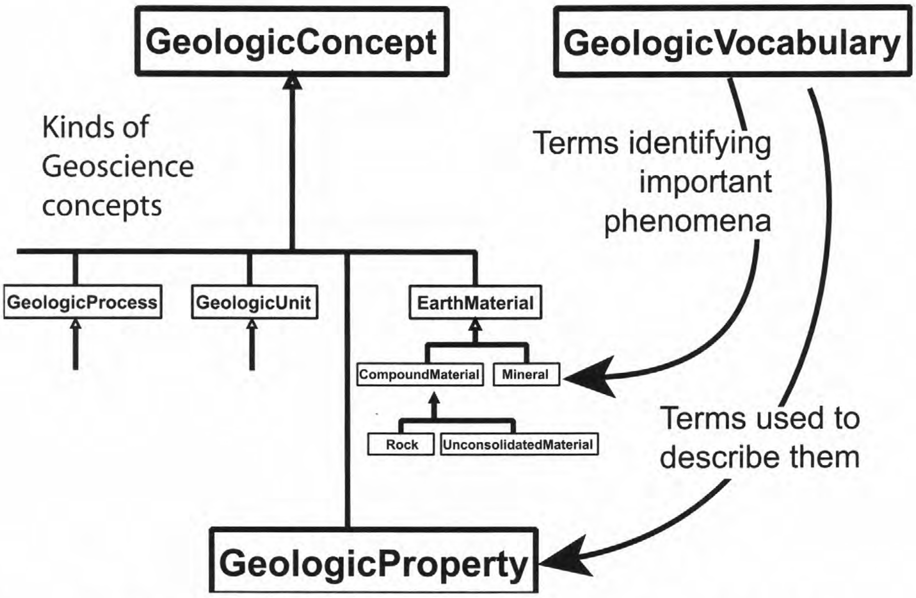


Figure 2. Relationship between NADM-C1 conceptual model and science language. GeologicVocabulary is a model element representing a controlled vocabulary or terminology system that contains the words used to enumerate important phenomena and to specify property values in formal descriptions; for example, a lithology lexicon is a collection of terms (science language, vocabulary) that correspond to different kinds of compound Earth materials.

> ish gray. It [grain size] ranges from aphanitic to coarse-grained, and the coarser grained diabase has ophitic texture." (Peterson, 1962)

In this sentence, the term diabase identifies a rock type that must be defined in some controlled vocabulary if it is to convey a precise meaning. The description here specifies three properties—color, grain-size, and fabric. The values for these properties is given using terms "dark gray" or "greenish gray" for color, "aphanitic" and "coarse-grained" specifying a grain-size range, and "ophitic" specifying a fabric or relationship between mineral grains in the rock. These terms would have to be defined in a controlled vocabulary in order to have unambiguous meaning in a database.

DESIGN OF NGMDB TERMINOLOGY

The NGMDB terminology system design consists of three parts. First is the actual list of terms that encompass the range of concepts that will be represented in the information system. Second is a specification of how descriptions are constructed to define concepts within the system. These description schema are defined by a conceptual data model that identifies the objects, properties, and relationships inherent in concepts to be represented in a database. The NGMDB science language uses description schema from the NADM-C1 (Data Model Design Team, 2003) model, along with some extensions to that model. For example, the description of the thickness of a lithostratigraphic unit includes the following properties: 1) reported minimum thickness; 2) reported maximum thickness; 3) typical thickness reported or interpreted; 4) source of typical thickness (field geologist, person who parsed the map legend, or other); 5) reported unit of measure; 6) original free text description or, for new data, a summary text description; 7) comments and/or processing notes from the person who parsed the legend; and 8) reference to a spatial extent (such as a rock body, borehole, outcrop, some part of a quadrangle). The description scheme also specifies the cardinality for each attribute (required, optional). The third part of the terminology system is the collection of definitions for terms included in the system.

The NGMDB terminology will attempt to follow the criteria of non-vagueness, non-ambiguity, and non-redundancy (de Keizer and Abu-Hanna, 2000). Non-vagueness prescribes that definitions of vocabulary terms must be sufficient to identify a particular concept. Non-ambiguity prescribes that definitions must address exactly one term. Non-redundancy specifies that a description has only one corresponding most-specific term. Because the NGMDB is intended to be useful to a wide variety of users, definitions of terms must be intelligible to non-geologists. In short, definitions of terms need to be clear, unambiguous, and non-overlapping.

The NGMDB science language is a concept vocabulary. The identity of a controlled concept (represented by a term from the vocabulary) is based on its definition, not the actual word used. The preferred name or term associated with the concept in the vocabulary is a label for the concept used in human communication. Each controlled concept has these properties:

- Unique identifier
- Name (a term)
- Definition
- Tracking (information on source of term)

Hierarchical arrangement of concepts is fundamental to many aspects of human thinking (Murphy and Lassaline, 1997). Most of the concepts that geologists use have a hierarchical structure, from very general to very specific; and for any observation or interpretation, terms consistent with the geologist's level of confidence are used. Hierarchical relationships between complex concepts are non-unique. For example, bedding partings in a sedimentary rock are a kind of planar structure and a kind of sedimentary structure. Multiple hierarchies may be defined, each valid and useful in some context. Implementation of the science language should include support for multiple explicitly defined hierarchies (Brodaric and others, 2002). For terms that have formal definitions constructed in the information system, description logic inference tools can determine hierarchy based on subsumption relationships between the normative descriptions.

Terms

Terms (controlled concepts) within a controlled vocabulary may be *atomic* or *analyzable*. The atomic terms are irreducible in the database environment; they are either numbers, or terms defined by narrative text. The text definition is meant to convey the meaning of a term to a human user, but is not amenable to automated analysis other than a simple string or number comparison of the identifier for the term. In the example given above, atomic terms specify color (dark gray) and fabric (ophitic). Another example would be terms used to describe outcrop character, such as "bouldery" or "ledgy".

Analyzable terms are specified by a linked collection of properties and relationships according to some formal description structure (description schema, see below). Analyzable terms may be compared on the basis of the values of the properties in their formal description. In the above example from Peterson (1962), the grain-size property is specified using analyzable terms "aphanitic" and "coarse-grained". These terms correspond to some quantitative range of particle diameters in the rock, commonly specified by a minimum and maximum diameter for the size-range represented by each term. The description associated with an analyzable grain size term might thus consist of a minimum diameter property and a maximum diameter property, each specified by a length property. The length property might in turn be specified by a description that includes a quantity property and a measurement unit property. The atomic concepts in this description would be the numbers that specify quantity, and terms that specify measurement systems (millimeters, inches, etc..).

The atomic terms in a vocabulary may be conceptually simple, but in many cases, atomic terms may represent complex concepts that could be represented by description schema. The choice of when to model and implement the description of a complex concept or to represent the concept using an atomic term is based on the usefulness of having an analyzable description. Implementation of formal descriptions for terminology associated with some concept allows users to develop alternate terminology, or more specific terminology, for the concept that may be integrated with other terminology systems that use the same description scheme for that concept. The granularity (level of detail) of terms in a vocabulary must be considered in the design of a practical terminology system. A vocabulary consisting only of atomic terms is relatively simple to implement, but is difficult to extend and customize. Such a vocabulary would probably be limited to more general terminology because of the difficulty in defining specialized terminology in such a way that it will gain acceptance. All hierarchical relationships would have to be explicitly represented. Processing queries in such a system would be relatively simple, and could probably be done using Structured Query Language(SQL). Systems in which formal descriptions are constructed for terms would allow users to customize terminology without losing the ability to integrate their data with others. Hierarchical relationships based on subsumption could be determined using description-logic inference software. The descriptions would still have to be based on a common terminology, but at a conceptually simpler level, for which wide agreement on definition is easier to attain. The cost is that such systems are more difficult to implement, and the query processing might require logical analysis of descriptions well beyond the capability of SQL databases, necessitating a logic-based inference processor.

Description Schema

A description schema is a specification of the properties associated with a concept, and its relation to other concepts. Particular instances of the concept are characterized by particular values for these properties and relationships. The NADM-C1 model includes description schemes for major geoscience concepts: materials, geologic units, and structures. Figure 3 shows a simplified version of the NADM-C1 description scheme for EarthMaterial. This scheme is the template for description of

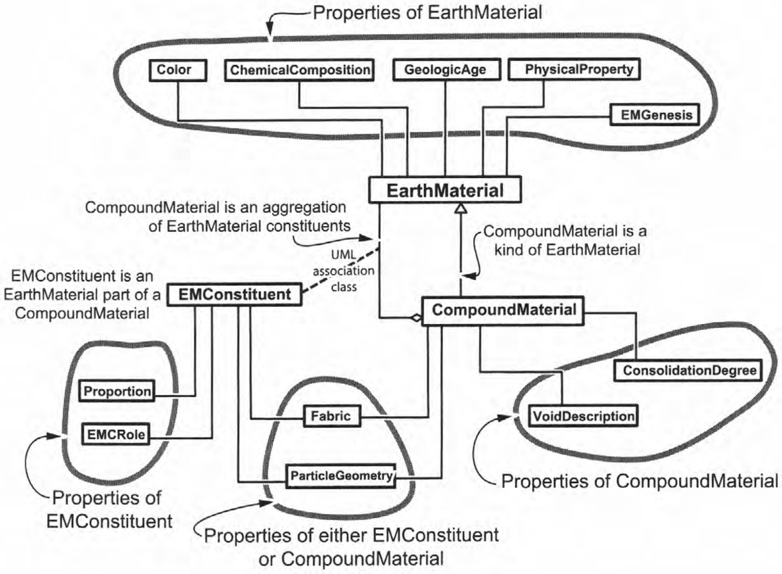


Figure 3. Schema for description of EarthMaterial from NADM-C1. UML notation (Rumbaugh and others, 1999) is used; multiplicities for association ends are not shown. EMConstituent is a UML association class, which represents a link between objects in which each link instance has individual attributes. Circles drawn with thick gray lines group properties that have the same range.

any material component of the Earth. An abstract class named CompoundMaterial represents materials that are aggregates of particles, and includes rocks and unconsolidated material. The scheme allows any material to be described by color, chemical composition, age, genesis (geologic history), or physical property attributes. A compound material (a rock) is described as an aggregation of constituent parts, each of which may be any Earth material. Some attributes are specific to compound materials, such as the degree of consolidation and description of the void spaces between particles. Some attributes are specific to constituents (an occurrence of an Earth material as part of a compound material), namely the proportion that the constituent forms in the aggregation, and a role property that specifies the relationship of the constituent to the whole aggregate material (for example, phenocryst, groundmass, matrix, framework, etc.). Finally, there are attributes that may apply to either individual constituents or to the aggregate as a whole, such as fabric and particle geometry. For instance, plagioclase may be a constituent as phenocrysts, with a grain size (part of particle geometry, see below) of 2-5 mm. Alternatively, a particular sandstone (compound material) may be described as fine-grained (a grain size attribute, part of the particle geometry group of properties).

Properties may be specified using atomic terms from a geologic vocabulary. For instance, consolidation degree in figure 3 may have an associated hierarchical vocabulary of atomic terms (fig. 4). Other properties may themselves have analyzable structure, with descriptions of property values in terms of other (generally simpler) properties. In this case, a vocabulary may be thought of as a collection of property descriptions, with a corresponding term for each description. The terms from this vocabulary effectively specify property values from the corresponding description.

As a demonstration of the relationship between an analyzable property description and a geologic vocabulary, figure 5 shows a detail of particle size description. Particle size is a component in the description of particle geometry, a property of either a compound material or a material constituent (fig. 3). Particle size may be specified quantitatively in a variety of ways. In figure 5, the simple approach of specifying a maximum and minimum grain diameter is the description scheme for a quantitative particle size description (QuantitativePSD). Each diameter is in turn a measured quantity, for which the description scheme specifies a numeric quantity and a measurement unit (determined by an atomic term from a vocabulary of measurement unit systems). The table in the lower left of the figure provides an example of four quantitative particle size descriptions. Several terminology systems have been proposed for specifying particle size in rocks. Terms from one of these systems may be used to populate a qualitative (terminological) particle size description (QualitativePSD in fig. 5). The table in the lower right of figure 5 includes several terms from one vocabulary for particle size (Wentworth, 1922). Each of these terms has a corresponding quantitative particle size description; thus they are analyzable terms.

APPLICATION OF NGMDB TERMINOLOGY

The National Geologic Map Database is envisioned as a distributed, federated information system, with nodes controlled by state geological surveys, the U. S. Geological Survey, and perhaps by interested business and academic collaborators. Sheth and Larson (1990) discuss the general framework and some approaches to such information systems. The NGMDB system will be distributed be-

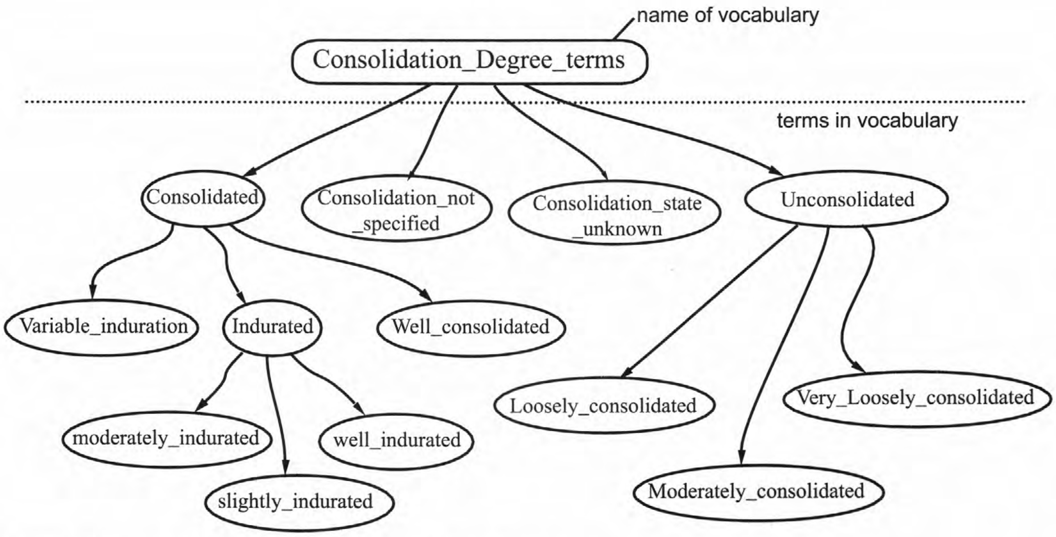
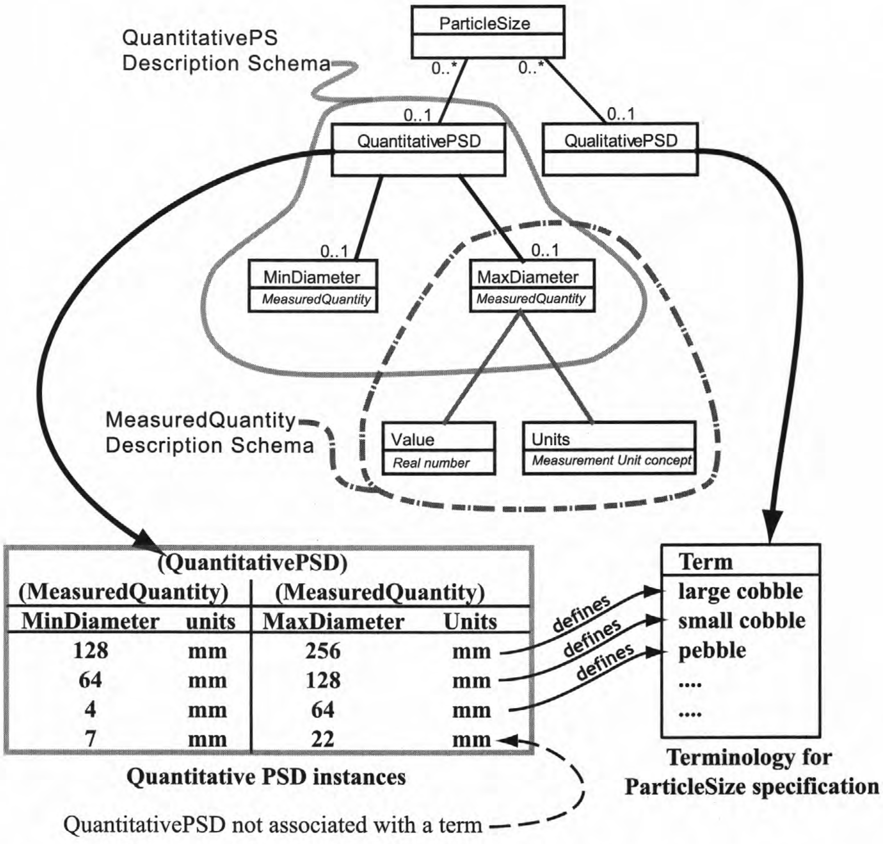


Figure 4. Vocabulary for consolidation degree property.



(QuantitativePSD)			
(MeasuredQuantity)		(MeasuredQuantity)	
MinDiameter	units	MaxDiameter	Units
128	mm	256	mm
64	mm	128	mm
4	mm	64	mm
7	mm	22	mm

Figure 5. Particle size description. PSD is an abbreviation for 'ParticleSizeDescription'.

cause data are created and managed by different agencies. The system is federated in that all nodes on the system will interoperate using a common conceptual model and a shared terminology. Individual nodes may implement the conceptual model and standard terminology directly, or may implement something different and interoperate with the network using a software transformation between their local data model and the common, conceptual data model. The common conceptual model will include NADM-C1, with extensions for description of geologic units and inclusion of feature level metadata.

Because the unifying feature of the NGMDB is the common conceptual model and shared terminology, to function as a distributed system each agency's database must conform to, or interoperate with, the common conceptual model and shared terminology. The simplest route to NGMDB conformance is to use the standard science vocabularies. If desired, users can modify the NGMDB standard vocabulary or build their own vocabularies. In order to be considered NGMDB-conformant, user-defined vocabularies likely will need to:

1. Define the scope of their vocabulary or (terminology)
2. Define terms in a human-intelligible form (text)
3. Build formal definitions rooted in the standard science vocabulary, using description schema from the common conceptual model, and a standard, well documented syntax. A standardized format is necessary but has not yet been determined.
4. Provide a correlation of each term in the vocabulary with the most specific subsuming term from the standard vocabulary.

The NGMDB Science Language is now under development, and will be made publicly available to facilitate implementation of conformant databases. The NGMDB terminology is intended to include:

1. Vocabularies for all geologic properties identified in the NADM-C1 data model
2. Hierarchical vocabularies for classifying rocks and unconsolidated materials, geologic structures, and geologic processes (EarthMaterial, GeologicStructure and GeologicProcess from NADM-C1)
3. Terminology required for the NGMDB relational database implementation of NADM-C1 (for example, vocabulary for measurement units)
4. Terminology for additional phenomena, properties, and metadata deemed necessary by collaborators in NGMDB prototype projects (such as engineering properties).

Figure 6 shows the framework for some of the vocabularies that eventually will be needed for the National Geologic Map Database. Clearly, development of these vocabularies is an ambitious goal that will take some time to achieve. Fortunately, other efforts are underway that share the same objectives, particularly the science language activities of the North American Data Model Steering Committee Science Language Technical Team (SLTT), this volume; <http://geology.usgs.gov/dm/steering/teams/language/charter.shtml>, and the Cyberinfrastructure For The Geosciences (GEON) project, a large, National Science Foundation-funded effort to develop cyberinfrastructure for geological sciences. Development of NGMDB science language is thoroughly integrated with science language development by the SLTT. Draft language developed by the SLTT is being incorporated directly into the draft NGMDB science language vocabularies. The National Geologic Map database project is actively exploring avenues of collaboration with GEON to accelerate development of formal geologic vocabularies for use in the construction of geologic information systems.

REFERENCES

Blatt, H., Middleton, G.V., and Murray, R., 1980, Origin of sedimentary rocks (2nd ed.): Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 782 p.

Boggs, Sam, Jr., 2001, Principles of sedimentology and stratigraphy (3rd ed.): Upper Saddle River, New Jersey, Prentice-Hall, Inc., 726 p.

Brodaric, B. M., Boisvert, E, and Patera, A., 2002, A set-theoretic technique and software for managing multiple-classification in geological hierarchies, *in* Proceedings, International Association of Mathematical Geologists: annual meeting.

Brodaric, Boyan, and Gahegan, M., 2000, Geoscience map data models, Open systems GIS and semantics, *in* Proceedings, GeoCanada2000-The Millennium Geoscience Summit, Calgary, Alberta, p. 7.

Brodaric, Boyan, and Hastings, Jordan, 2002, An object model for geologic map information, *in* Richardson, D., and van Oosterom, P., eds., Advances in Spatial Data Handling, 10th International Symposium on Spatial Data Handling: Heidelberg, Germany, Springer-Verlag, 562 p.

Brodaric, Boyan, Boisvert, Eric, Patera, Antonio, 2002, A set-theoretic technique and software for managing multiple-classification in geological hierarchies: 2002 Annual Conference of the International Association for Mathematical Geology, Berlin.

Collinson, J.D., and Thompson, D.B., 1989, Sedimentary structures (2nd ed.): London, Unwin Hyman, 207 p.

Data Model Design Team, 2003, NADM conceptual model 1.0, A conceptual model for geologic map information: North American Data Model Steering Committee, available at http://geology.usgs.gov/dm/steering/teams/design/NADM-C1.0/NADMC1_0.pdf (accessed 02/04/2004), 55 pages.

de Keizer, N.F., and Abu-Hanna, A., 2000, Understanding terminological systems II: Experience with conceptual and formal representation of structure: Methods of Information in Medicine, v. 39, p. 22-29.

de Keizer, N.F., Abu-Hanna, A., and Zwetsloot-Schonk, J.H.M., 2000, Understanding terminological systems I - Terminology and typology: Methods of Information in Medicine, v. 39, p. 16-21.

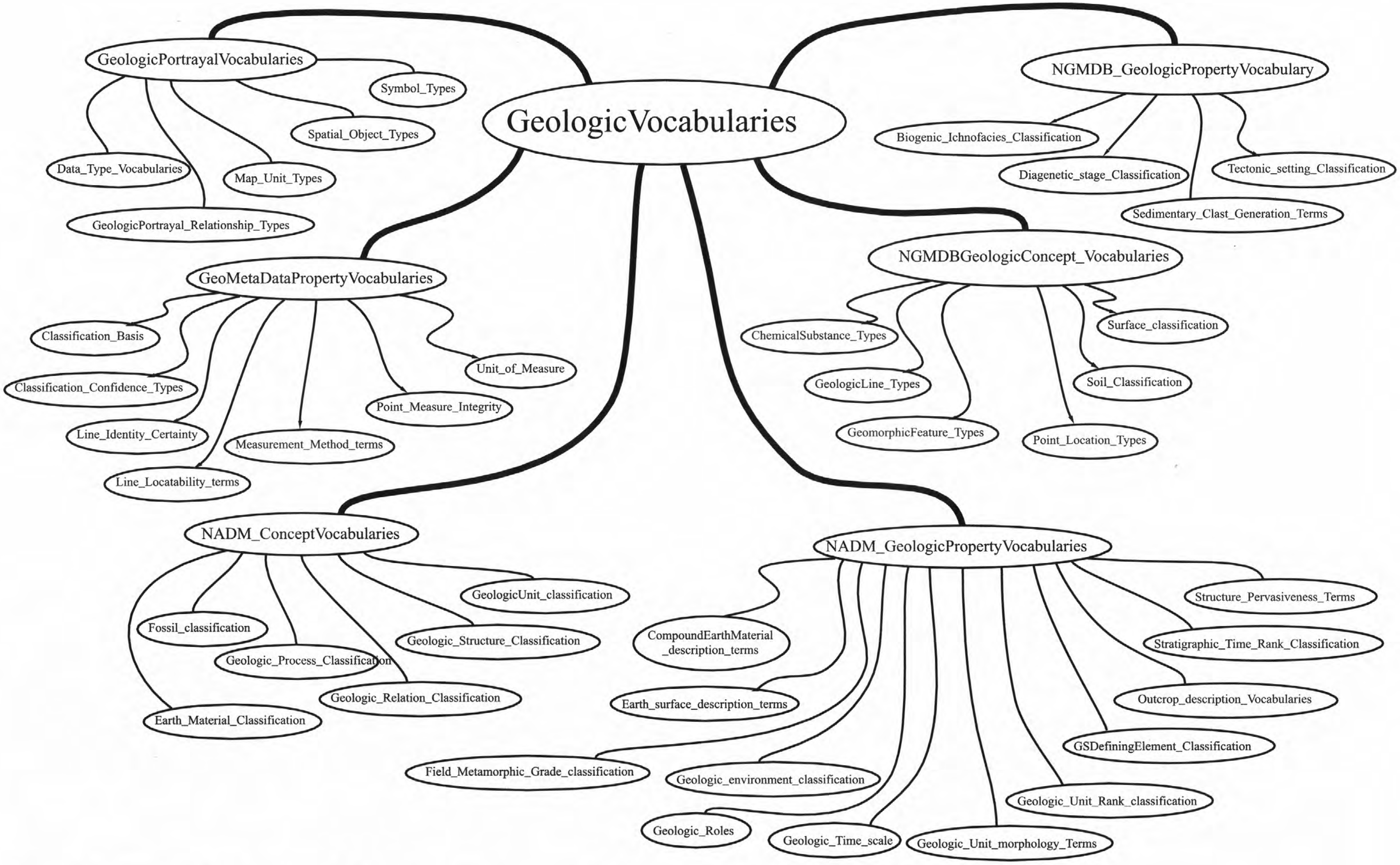


Figure 6. Framework of geologic vocabularies that will be important for the National Geologic Map Database.

Folk, R.L., 1968, Petrology of sedimentary rocks: Austin, Texas, Hemphill's Bookstore, 170 p.

Giles, J.R.A., Lowe, D.J., Bain, K.A., 1997: Geological dictionaries--critical elements of every geological database: Computers & Geosciences, v. 23, 621-626.

Murphy, G. L., and Lassaline, M. E., 1997, Hierarchical structure in concepts and the basic level of categorization, *in* Shanks, David, and Lamberts, Koen, eds., Knowledge, concepts, and categories: Cambridge, Ma., The MIT Press, p. 93-132.

North American Commission on Stratigraphic Nomenclature, 1983, North American stratigraphic code: American Association of Petroleum Geologists Bulletin, v. 67, no. 5, p. 841-875.

Peterson, N. P., 1962, Geology and ore deposits of the Globe-Miami District, Arizona: U. S. Geological Survey Professional Paper 342, 151 pages.

Pettijohn, F.J., 1957, Sedimentary rocks (2nd ed.): New York, Harper Brothers, 719 p.

Potter, P.E., Maynard, J.B., and Pryor, W.A., 1980, Sedimentology of shale: New York, Heidelberg, Berlin, Springer-Verlag, 306 p.

Richard, S. M., 1999, Geologic concept modeling, with examples for lithology and some other basic geoscience features, *in* Soller, D. R., Ed., Digital Mapping Techniques 1999, Workshop Proceedings, U. S. Geological Survey Open-File Report 99-386, p. 59-75, <http://pubs.usgs.gov/of/1999/of99-386/richard.html>.

Richard, S. M., in press, Geoscience concept models, *in* Sinha, K, and Baru, C, Geoinformatics: Geological Society of America Special Paper.

Rumbaugh, J., Jacobson, I., Booch, G., 1999, The unified modeling language reference manual: Reading, Ma., Addison-Wesley, 576 p.

Sheth, A.P., and Larson, J.A., 1990, Federated database systems for managing distributed, heterogeneous, and autonomous databases: ACM Computing Surveys, v. 22, p. 183-236.

Soller, D. R., Brodaric, Boyan, Hastings, J. T., Wahl, Ron, and Weisenfluh, G. A., 2002, The central Kentucky prototype: An object-oriented geologic map data model for the National Geologic Map Database: U. S. Geological Survey Open-File Report 02-202, 39 p., <http://pubs.usgs.gov/of/2002/of02-202/>.

Wentworth, C.K., 1922, A scale of grade and class terms for clastic sediments: Journal of Geology, v. 30, p. 377–392.

Williams, H., Turner, F.J., and Gilbert, C.M., 1954, Petrography: San Francisco, W.H. Freeman and Co., 406 p.

Geologic Map Database Implementation in the ESRI™ Geodatabase Environment

By Stephen M. Richard

Arizona Geological Survey
416 W. Congress # 100
Tucson, AZ 85701
Telephone: (520) 770-3500
Fax: (520) 770-3305
email: Steve.Richard@azgs.az.gov

INTRODUCTION

The Arizona Geological Survey (AZGS) has been producing 1:24,000-scale geologic map deliverables for the National Cooperative Geologic Mapping Program since the inception of the program. Digital compilation and production of map deliverables started on a prototype basis in 1998, using ArcEdit for map digitizing, and ArcView 3.2 for cartographic production. Map layouts were converted into Adobe Acrobat documents, and are available on CDROM or as hard copy generated using an HP 755CM large-format inkjet printer. The release of version 8.3 of ArcGIS, with the capability of enforcing topological relationships between geologic lines (faults and contacts) and polygons, makes it possible to do the entire map compilation and production process in one software environment. Digital compilation and cartography is done by the mapping geologists at the AZGS, and the ability to do all map production work in one software environment means the geologists need only learn one software package. In order to reduce the amount of software training necessary, we decided to convert to map production using ArcGIS 8.3, and in 2002-2003 produced 7 quadrangle-map deliverables using this product.

AZGS geologic map databases have been produced with spatial data in ArcInfo coverages and with thematic data (mostly metadata, so far) in Microsoft Access databases. The ESRI Geodatabase data model used in ArcGIS 8.3 allows implementation of an integrated database containing both spatial data and thematic data in one package. ArcGIS 8.3 introduced a new model for representing topologic relationships between geologic contacts and faults and the polygons representing outcrop of geologic units. The new software release also added some subtype and domain definition functionality. Production of geologic maps using the new software allowed us to investigate these capabilities for improved geologic map representation.

DATABASE CONCEPTUAL DESIGN

The data structure described here builds on the design described in Richard and Orr (2001), and is intended to conform to the conceptual model produced by the Data Model Design Team of the North American Data Model Steering Committee (NADMSC, 2003) (http://geology.usgs.gov/dm/steering/teams/design/NADM-C1.0/NADMC1_0.pdf).

GEODATABASE IMPLEMENTATION

The geologic map database implementation consists of a feature dataset (see table 1) that contains geologic points, lines, and polygons feature classes; a feature dataset for cross-sections; a feature dataset for cartographic objects in the map layout that have fill symbols corresponding to geologic units (map explanation, correlation of map units); and a group of tables containing thematic data. The AZGS implementation uses personal geodatabases, which are Microsoft Access .mdb files.

The geologic spatial data is in a feature dataset named GeologyFeatures. The spatial reference for the geology feature dataset was defined using a standard spatial reference system designed for the state of Arizona, with meters as the unit of measurement (Table 2). The coordinate system provides better than 1 millimeter (mm) possible precision for horizontal coordinates (see scale value in Table 2, 1/1533.9 m), which is considered quite adequate given that location uncertainty assigned to mapped geologic contacts by AZGS geologists is generally no better than 10 m.

The GeologyFeatures dataset contains a point feature class, line feature class and polygon feature class; and one topology. Points represent stations occupied in the field; they may be associated with one or more structural measurements, samples, photos, rock descriptions, etc. The line feature class represents contacts (depositional,

Table 1. Definition of selected ESRI geodatabase terms.

Term	Definition
Feature dataset	A collection of related feature classes, all of which share a common spatial reference system.
Feature class	A collection of spatial objects with their attributes.
Topology	A collection of rules concerning spatial relationships between feature classes in a single feature dataset.
Spatial reference	Specification of a coordinate system, spatial domain (minimum and maximum X, Y, and Z coordinates), and precision.
Subtype	Subset of data instances in a feature class or table, differentiated by integer values in a database field of that class or table.

Table 2. Spatial domain for Arizona (meters).

Coordinate	Min	Max	Scale
X	-200000.0	1200000.0	1533.92
Y	3000000.	4400000	1533.92
Z	-50000	10000	35791.39

gradational, intrusive…) and faults, and the polygon feature class represents outcrop of geologic map units on the map horizon depicted. User-defined fields for the line and polygon geology feature classes are identical to those described in Richard and Orr (2001). The table associated with point locations has been modified from the GeoPnt table described in Richard and Orr (2001) using feedback from AZGS geologists. The Station data feature class attributes (Table 3) are designed to accommodate all standard data collected in the field. Multiple instances of this class may share the same location, and contain different information 'bits' collected at that site. One topology is defined with the lines and polygons participating. Rules for this topology are summarized in table 4. These rules implement the same topology implemented in an ESRI polygon coverage.

Subtypes

The geodatabase data model allows one field in a table or feature class to be specified as a subtype field. The field must contain integer values. This allows subsets of a feature class to have distinct value domains defined for database fields, and to have distinct topology rules. The subtype hierarchy is only one level deep, and only one hierarchy may be represented. The subtype values must be entered by hand in a dialog, along with any description of the subtype (fig. 1).

Several fields might be used as a subtype field for geologic lines (table 5). After consideration of the pros and cons, the CartoObjID field (symbolization selector field) was chosen to define subtypes. This has proven very useful in several ways. During data entry, the symbol is selected before digitizing a feature, and is immediately displayed on screen using standard geologic symbolization (defined using a standard legend file). The selected symbol provides constraints on data that may be entered in other fields. For instance, selection of a solid line implies a maximum value for the Accuracy field, depending on the map accuracy standard. The symbol also implies one or a few possible ConceptID values; if a thrust fault symbol subtype is selected, then the ConceptID must be the value for the thrust fault concept.

Domains

The geodatabase allows different value domains to be specified for fields in a table based on the subtype code. These domains are presented to the user as pick-lists (combo boxes) in the user interface (fig. 2). Definition of value domains is similar to definition of subtypes—entries must be typed value by value in a dialog box. There is no facility for representing hierarchical value domains. Construction of the domain lists is facilitated by a tool included in the ArcObjects developer kit (the extension is named Table2Domain) that constructs a domain from values from a stand-alone table. Each domain required must be defined as a separate table.

Relationships

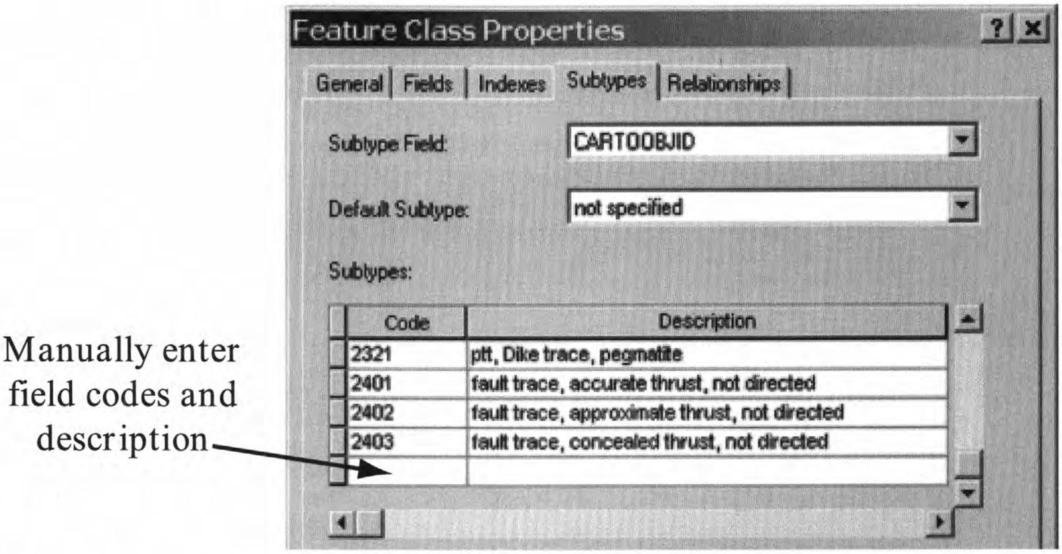
The geodatabase model includes provision for repre-

Table 3. Fields for station data feature class

Name	Description
OBJECTID	Long Integer, Internal geodatabase field. Unique identifier for each station record feature. In personal geodatabase, this is a Microsoft Access autonumber field.
Shape	Internal geodatabase field. Representation of feature geometry.
DataSetID	Long Integer. Unique identifier for this dataset (object class in geodatabase parlance). Analogous to ID field in GDB_ObjectClasses table in geodatabase .mdb file. Because the ObjectClass interface does not allow the user to assign object class or data set identifies, these must be maintained externally (AZGS DataSetAz table, see Richard and Orr (2001)).
STATION	Text, width 32. Field geologist's unique identifier for station. Multiple station records may be associated.
GROUPING	Text, width 64. Text string that may be used to group data observations. Example: group measurement records for a foliation orientation and a lineation orientation measurement for which the foliation contains the lineation to form a compound fabric. Informal field for use by field geologist.
HEADING	Text, width 32. Text string that the geologist may use to classify data observations to facilitate analysis. For instance, geologist may have informal map unit designations, or an alternate system of classifying structural data (S_1, S_2, S_3...). Informal field for use by field geologist.
DESCRIPTION	Text, width 255. Text description of a geologic feature at the station. The Grouping and Heading fields allow the geologist to break up and classify notes taken at a station into different records according to the topic. For instance lithology descriptions might all be in records with a header like 'lithology', and alteration descriptions might all have a heading 'alteration'. Informal field for use by field geologist.
STRIKE	Decimal number. Azimuth or bearing of orientation measurement, from north, in degrees. For surfaces, use right-hand rule—when facing in azimuth direction, surface dips to the right. For overturned surfaces, the dip value is reported >90, and the strike azimuth is reported in the direction for which the surface is tilted down on the right, through vertical to attain overturned disposition. For directed linear features, such as mylonitic lineation with known sense of shear, azimuth is reported in positive direction, and plunge (dip) is <0 if the positive direction is up-plunge. Domain: -360 < Strike < 360.
ERR_STRIKE	Decimal number. Angular uncertainty in determination of orientation azimuth; for example orientation = Strike +/- Err_strike. Analogous to alpha95 value in spherical statistics. Domain 0 < Err_strike < 180.
DIP	Decimal number. Plunge of linear feature or dip of planar feature. May be >90 for overturned surfaces, or <0 for directed linear features that have an up-plunge positive direction. Domain –90 < Dip < 360.
SAMPLE	Text, width 24. Geologist's identifier for rock sample collected at station. Informal field for use by field geologist.
ANALYTICAL	Text, width 24. Sample collection purpose, if sample collected for analysis. Informal field for use by field geologist.
Label	Text, width 24. Text string for labeling the point location. Examples: dip or plunge value to display with structure symbol, or Sample identifier string.
TYPE	Text, width 24. Text string, name of concept specified by {TypeID, TypeDS}. This is redundant, but provided for ease of use by geologists. Domain: Names of concepts in dataset(s) specified by TypeDS field value.
TypeID/TypeDS	Compound foreign key, link to classification concept that identifies the kind of of station data recorded in this record. May be notes, sample, unclassified, structural measurement, etc. For structural data, identifies the kind of structure orientation data described. It is a foreign key that joins to the ConceptID field of the ClassificationConcept table.
SYMBOL	Text, width 24. Text string, name of symbol specified by {CartoObjID/CartoObjDS}. This is redundant, but provided for ease of use by geologists. Domain: Names of symbols in dataset(s) specified by CartoObjDS field value.
CartoObjID/CartoObjDS	Compound foreign key, link to the symbol used to depict a feature in the default map visualization. It is a foreign key that joins to the CartoObjID field of the CartographicObject table.
ROTATION	Decimal number. Rotation value for oriented structure symbol, specified using ArcView 3.2 rotation conventions (mathematical convention)—0 is due east, positive rotation is counterclockwise.
UTME	Decimal number. UTM easting coordinate for station location. Duplicates information contained in shape field, but provided for transportability to other GIS software.
UTMN	Decimal number. UTM northing coordinate for station location. Duplicates information contained in shape field, but provided for transportability to other GIS software.
ACCURACY	Decimal number. The spatial uncertainty in the location of a feature, in meters. For example, a value of 10 for a point feature indicates that the geologic entity represented by the point is within 10 meters of the reported coordinates. Domain: >0.
mapHorizon	Text, width 128. Text string identifying the surface (generally earth surface) upon which the mapped outcrop traces are located. This can be thought of as a proxy for Z (elevation) values for features.
TrackingNotes	Notes on origin of feature; used during dataset construction to build complete tracking record identified by {TrackingID, TrackingDS} link.
TrackingID/TrackingDS	Compound foreign key, link to the origin tracking (metadata) for each object. It is a foreign key that joins to the TrackingID field of the TrackingRecord table.

Table 4. Topology rules for geologic lines and polygons

1. Lines may not intersect.
2. Lines may not have dangles.
3. Polygon boundaries must have coincident geologic lines.
4. Polygons must not overlap.
5. Polygon may not have gaps.



Subtype Field: CARTOOBJID

Default Subtype: not specified

Subtypes:

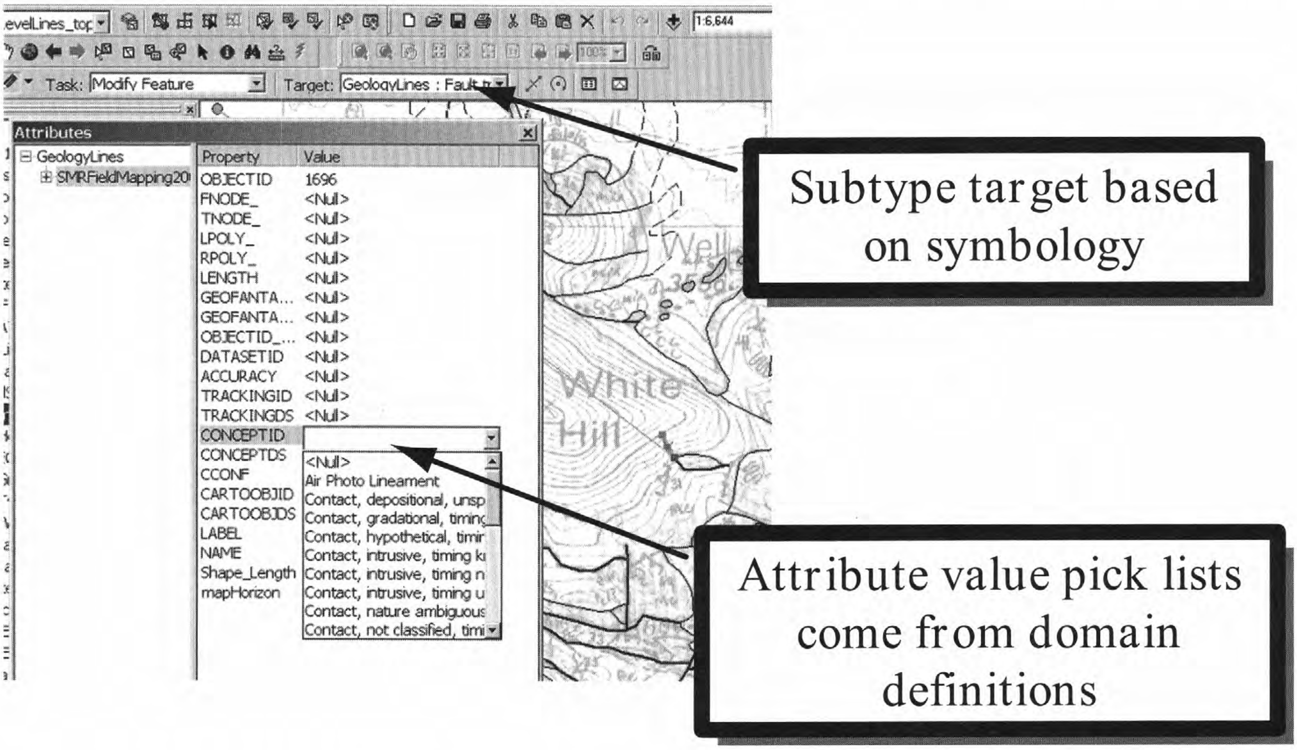
Code	Description
2321	ptt, Dike trace, pegmatite
2401	fault trace, accurate thrust, not directed
2402	fault trace, approximate thrust, not directed
2403	fault trace, concealed thrust, not directed

Manually enter field codes and description

Figure 1. Subtype definition dialog. CartoObjID codes are those defined by AZGS.

Table 5. Possible fields to use for determining geologic line subtype.

Field	Rationale	Pro	Con
ConceptID	Subtype based on nature of geologic feature represented.	Provides a good representation of science-based constraints, for example, faults may have dangles, but contacts may not have dangles.	Classification of line types is hierarchical with depth >1; can't be represented with geodatabase subtype implementation.
MapHorizon (would have to convert text to integer)	Base on surface that is host for geologic lines depicted on map.	Define separate topology for each map horizon (use geologic line rules). Inferred bedrock surface geology would not interfere with surficial geology. Avoid having a collection of identical features (geoLines, geoPolys) for each map horizon.	Not necessary in many databases.
CartoObjID	Symbolization of line—related to geologic line type, location uncertainty, map horizon.	Can assign symbolization during digitizing process. Other attribute values implied by symbol assignment can be inserted in table.	Not what is needed for analysis. Need to do considerable customization to take advantage of linkage to other fields.



Task: Modify Feature

Target: GeologyLines : Fault

Property	Value
OBJECTID	1696
FNODE_	<Null>
TNODE_	<Null>
LPOLY_	<Null>
RPOLY_	<Null>
LENGTH	<Null>
GEOFANTA...	<Null>
GEOFANTA...	<Null>
OBJECTID_...	<Null>
DATASETID	<Null>
ACCURACY	<Null>
TRACKINGID	<Null>
TRACKINGDS	<Null>
CONCEPTID	
CONCEPTDS	
CCONF	
CARTOOBJID	
CARTOOBJDS	
LABEL	
NAME	
Shape_Length	
mapHorizon	

<Null>
Air Photo Lineament
Contact, depositional, unsp
Contact, gradational, timinc
Contact, hypothetical, timir
Contact, intrusive, timing ki
Contact, intrusive, timing n
Contact, intrusive, timing u
Contact, nature ambiguous
Contact, not classified, timi

Subtype target based on symbology

Attribute value pick lists come from domain definitions

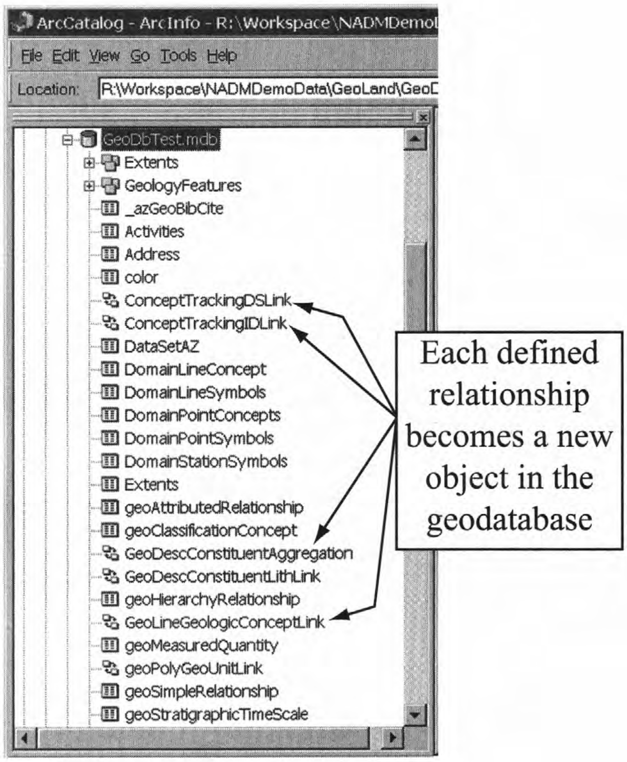
Figure 2. Subtype selection determines domains for some attributes. The Target control on the tool bar allows selection of the subtype for a new feature when digitizing. In this figure, assignment of a subtype value (fault) determines the value domain for the ConceptID attribute.

senting relationships as relationship classes. These may be either simple or complex (geodatabase parlance). Simple relationships are essentially relational table joins stored in the geodatabase. Complex relationships are represented by a correlation table in the geodatabase, allowing each relationship link to have associated properties. Both kinds of relationships appear in the geodatabase table of contents viewed in ArcCatalog (fig. 3). The ArcMap interface allows editing of data in related tables by selecting a source object in the relationship. Thus, a line or polygon may be selected, and properties specified in related tables may be edited. The Identify dialog (opened by clicking on a feature using the 'Identify' tool on the 'Tools' toolbar) allows for browsing of links to related tables, to inspect the contents of linked records. Unfortunately, there is no way to control which linked information is displayed (fig. 4). Unlike explicit joins, which may also be constructed in the ArcMap environment, fields in tables linked by a relationship class do not show up in the Table view of a feature class, and are not available for constructing queries.

DIGITIZING WORKFLOW AND MECHANICS

The transition from digitizing coverages using ArcEdit to digitizing features in a geodatabase required rethinking the map construction workflow. Issues included when and how to generate polygons from digitized lines, the best method of editing topologic features, and procedures for editing map data. Because maps are compiled from field sheets, the AZGS map compilation process is incremental. Polygons are constructed as line work is digitized from scanned, georeferenced field sheets. First, a group of lines is digitized from a field sheet. Then, these lines are planarized (similar to 'clean' in Arcedit—producing nodes at line intersections and snapping where necessary) by selecting the lines and clicking the 'Planarize Lines' button on the Topology tool bar. The topology validator is used to fix all dangles and undershoots. Faults that end within a polygon are legal dangles and are marked as exceptions. When the line topology problems are fixed, the lines defining polygons to be generated are selected, and the 'construct features' button on the Topology toolbar is clicked. The Target (Editor toolbar) must be set to the geology polygon layer, and the Task (Editor toolbar) to 'Create New Feature'. Once all polygons defined by line work on a field sheet scan have been constructed, a new field sheet scan may be started.

During the transition to map production in ArcGIS 8.3, some data were digitized using ArcEdit to produce coverages, which were then imported into geodatabase



ArcCatalog - ArcInfo - R:\Workspace\NADMDemo

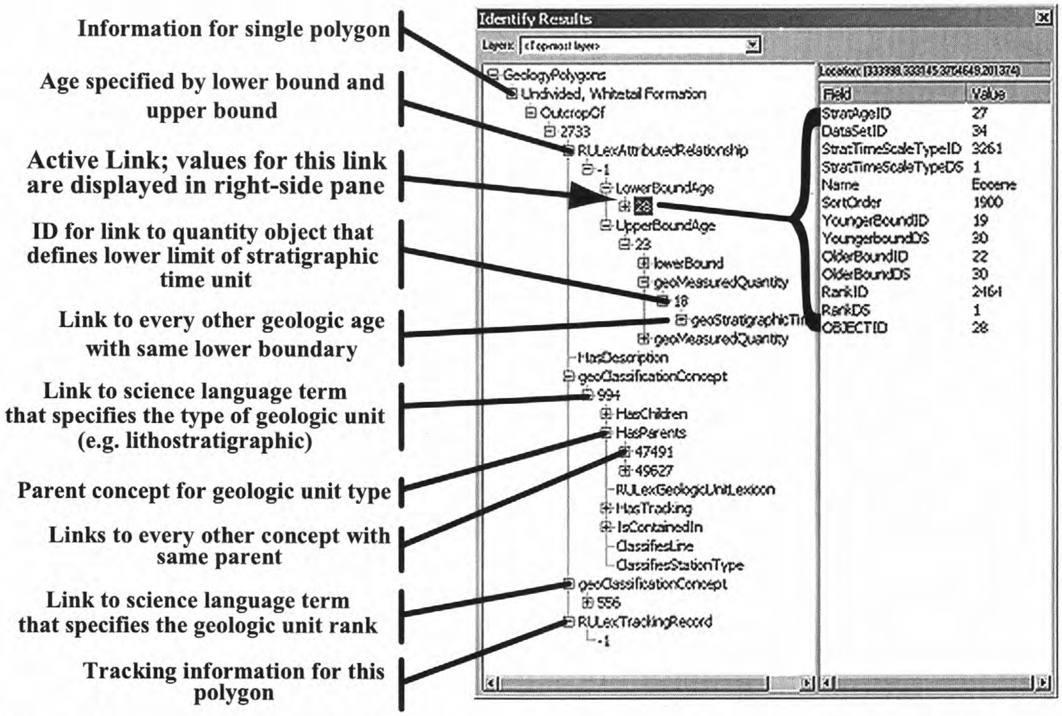
File Edit View Go Tools Help

Location: R:\Workspace\NADMDemoData\GeoLand\GeoD

```
GeoDbTest.mdb
  Extents
  GeologyFeatures
  _azGeoBibCite
  Activities
  Address
  color
  ConceptTrackingDSLink
  ConceptTrackingIDLink
  DataSetAZ
  DomainLineConcept
  DomainLineSymbols
  DomainPointConcepts
  DomainPointSymbols
  DomainStationSymbols
  Extents
  geoAttributedRelationship
  geoClassificationConcept
  GeoDescConstituentAggregation
  GeoDescConstituentLithLink
  geoHierarchyRelationship
  GeoLineGeologicConceptLink
  geoMeasuredQuantity
  geoPolyGeoUnitLink
  geoSimpleRelationship
  geoStratigraphicTimeScale
```

Each defined relationship becomes a new object in the geodatabase

Figure 3. Relationship classes in ArcCatalog table of contents.



Identify Results

Layers: <Top-most layer>

```
GeologyPolygons
  Undivided, Whitetail Formation
    OutcropOf
      2733
        RULexAttributedRelationship
          -1
            LowerBoundAge
              28
            UpperBoundAge
              23
                lowerBound
                geoMeasuredQuantity
                  18
                    geoStratigraphicTi
                geoMeasuredQuantity
        HasDescription
        geoClassificationConcept
          994
            HasChildren
            HasParents
              47491
              49627
            RULexGeologicUnitLexicon
            HasTracking
            IsContainedIn
            ClassifiesLine
            ClassifiesStationType
        geoClassificationConcept
          556
        RULexTrackingRecord
          -1
```

Location: (333938.333145 3764649.201374)

Field	Value
StratAgeID	27
DataSetID	34
StratTimeScaleTypeID	3261
StratTimeScaleTypeDS	1
Name	Eocene
SortOrder	1900
YoungerBoundID	19
YoungerboundDS	30
OlderBoundID	22
OlderBoundDS	30
RankID	2461
RankDS	1
OBJECTID	28

Information for single polygon

Age specified by lower bound and upper bound

Active Link; values for this link are displayed in right-side pane

ID for link to quantity object that defines lower limit of stratigraphic time unit

Link to every other geologic age with same lower boundary

Link to science language term that specifies the type of geologic unit (e.g. lithostratigraphic)

Parent concept for geologic unit type

Links to every other concept with same parent

Link to science language term that specifies the geologic unit rank

Tracking information for this polygon

Figure 4. Identify Results dialog for geologic unit information linked to a polygon. Without significant customization, the significance of the various links is not apparent, and much unnecessary information is presented. Age of unit has a lowerBoundAge and UpperBoundAge that are both links to stratigraphic ages from a geologic time scale (such as Eocene). Each of these stratigraphic ages has a lower and upper bounding age (in Ma) that are stored in linked measuredQuantity objects. Science language terms (geologic unit type and rank) are hierarchical, and the hierarchy links appear due to the defined geodatabase relationships, not only immediate parent and child terms, but also to every other term with the same parents.

feature classes. This process worked well using the 'Coverage to Geodatabase' wizard, accessible by right-clicking on the geodatabase icon in ArcCatalog. This produces both a line and a polygon feature class. In order to integrate these into a single topology, so that lines or polygons may be edited keeping polygon boundaries coincident with lines, a new topology must be defined that applies the rules in Table 4 to the new feature classes. The new line and polygon feature classes must be in the same feature dataset to create the topology. Care must be taken that the cluster tolerance defined for the topology is less than the fuzzy tolerance for the coverage; if the fuzzy tolerance is too small, the coverage lines must be generalized (at the workstation Arc command line) using a weed tolerance slightly greater than the cluster tolerance for the geodatabase topology.

The 'topology edit' tool (Topology toolbar) works well for editing polygon boundaries with coincident lines. When editing at the junction of three or more polygons must be done, it is commonly easier to use the standard edit tool (Edit toolbar) with snapping set to "on" between the line and polygon features. The snap environment—which feature classes will snap, and whether snapping will occur between nodes, vertices, or edges—is set using the 'Snapping…' choice on the Editor drop-down menu on the Editor tool bar. The snap tolerance is set via the 'Editing Options' dialog accessed through the 'Options' choice on the Editor drop-down menu on the Editor tool bar. A value of 3-5 meters (m) seems to work well for the Arizona spatial reference system. Setting the 'Sticky move tolerance' in the same dialog to a value of 5-10 pixels avoids frustration with features being selected and moved accidentally when trying to select nearby features.

Adding and removing polygons can be done in a variety of ways. The topology tools offer several features that can facilitate the process. A polygon may be created by splitting an existing polygon. Set the edit target to the polygon layer, and select the 'Cut polygon features' choice on the Task drop-down menu on the Editor tool bar. Select the polygon to split, and with the 'Sketch tool' (pencil button), draw the line where the polygon will be split. After splitting, validate the topology, and the new polygon boundary will show up as a violation of the 'Polygon boundaries must have coincident geologic lines' rule (Table 4). In fact there are two topology errors, one for each polygon, but both are fixed by creating a line along the boundary. Clicking on the topology error with the 'Fix topology error tool' (Topology toolbar) selects both errors. Right clicking with the 'Fix topology error tool' provides choices for fixing the error, and 'Create feature' is the obvious choice. Unfortunately, this will create two new, coincident lines, and a new topology error because of the rule 'Lines may not intersect'. Some trouble can be saved by having the 'Error Inspector' window open (click the 'Error Inspector' button on the Topology toolbar); then search for errors after doing the polygon split, right-click on one of the 'Lines may not intersect' errors and select 'Create feature'. This creates just one line, and finishes the split operation.

When the new polygon overlaps more than one existing polygon, the new polygon may be digitized on top of the existing polygons. Validating the topology at this point results in errors because of overlapping polygons, and a polygon boundary without coincident lines. Using the 'Fix topology error tool,' right click on the line error produced along the new polygon boundary and select 'Create feature' to create a line on the new polygon boundary. Select this line and all existing lines it intersects and click the 'Planarize Lines' button on the topology tool bar to split the existing lines where they intersect the new polygon boundary (and vice versa). Using the 'Fix topology error tool', right-click on the area error created by the new polygon (overlap with existing polygons) and select 'Create feature' to insert the new polygon into the map topology by intersecting it with the existing polygons. The map geometry must then be cleaned up by 1) merging the parts of the new polygon created where it overlapped different polygons in the original map; and 2) deleting lines that were polygon boundaries within the new polygon.

MAP UNIT DESCRIPTION

The basic elements of geologic unit description are shown in figure 5. The GeologicUnitLexicon is a vocabulary of geologic unit names. Each is associated with at least one description in the GeologicUnitDescription table that is called the normative description. Other descriptions in the GeologicUnitDescription table may be associated with a geologic unit. Each geologic unit description pertains to some body of rock represented by an Extent object. This may represent an outcrop (point location), the rock exposed in some area (one or more polygons), or the entire geologic unit (in the case of the normative description). The geologic unit description contains links to terms from a geologic vocabulary (NGMDBScienceLanguage) that specify various properties of the unit (links labeled 'PropertyTermLinks (many)' in Figure 5). Some properties may be specified by complex data objects that are collections of more specific properties and relationships. Some examples include geologic unit thickness, age, surface character, and genesis. A geologic unit is composed of one or more parts, represented by the aggregation relationship from GeoUnitDescription to GeoUnitPart in the top left of figure 5. The parts of a geologic unit may be 1) an Earth material specified by a term in a geologic lexicon of lithologic classes (NGMDBCompoundMaterialLexicon); 2) an Earth material specified by an Earth material description but not part of a vocabulary (CompoundMaterialDescription); 3) a geologic unit from a geologic unit lexicon (GeoUnitLexicon), for instance, in the description of a stratigraphic group, the parts are named (in a lexicon) formations; or 4) a geologic unit description

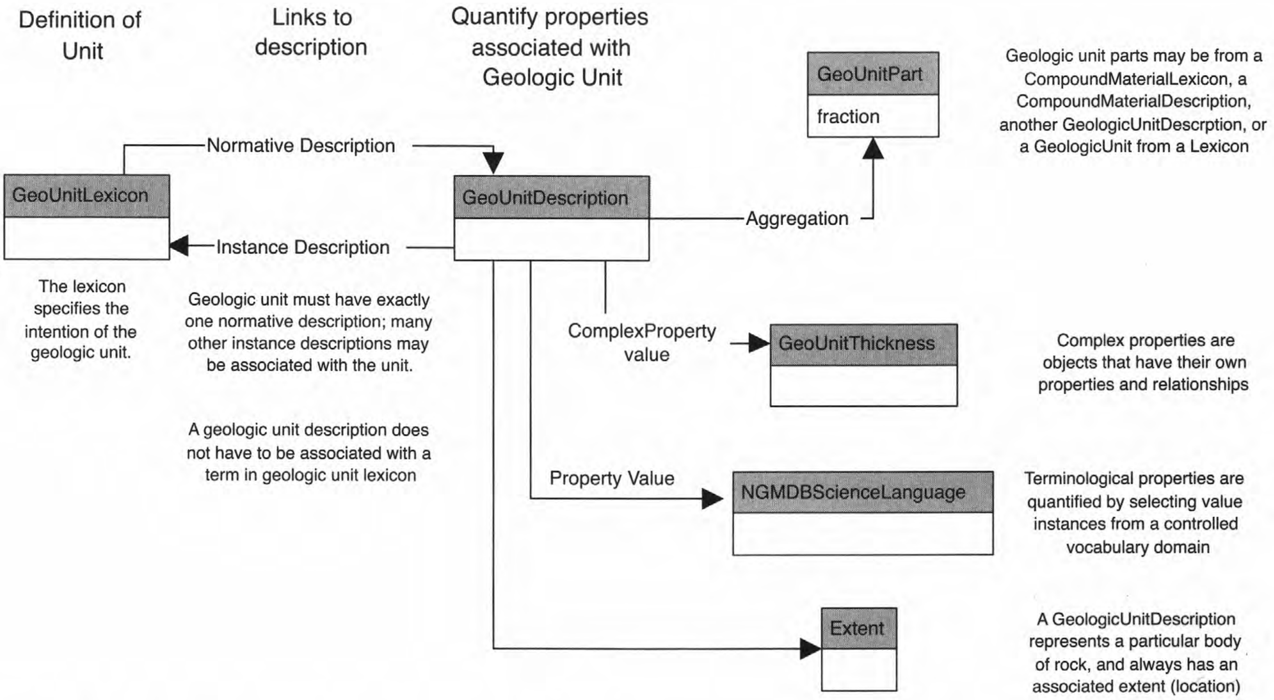


Figure 5. Generalized scheme for geologic unit description.

not part of a lexicon (GeoUnitDescription).

A more detailed entity-relationship diagram for the implementation of geologic unit description is shown in figure 6. A summary of the fields in each table is included in the appendix. Each table includes several standard fields. The Name field contains a text string used to identify each data instance. In the case of lexicon terms (GeologicUnit and CompoundMaterial) the name is the preferred name associated with a concept. For description tables, the name identifies the description, as an abbreviated summary of the description or some user-defined name. Tracking links (see below for discussion of 'ID' and 'DS' suffixes) associate data instances metadata documenting processing history and intellectual sourcing of data. OriginDate is an automatic field that records the data and time a record was created in the database.

Each geologic unit definition record in the GeologicUnitLexicon table contains a link to a type and rank term in the NGMDBScienceLanguage (GeoUnitTypeID, RankID); a name for the unit; a free text description explaining the intent of the unit; and the URL of the GeoLex record for the unit, if one exists. The geologic unit definition also includes a link to a normative description (DescID/DescDS), and a link to a geologic age description (GeolAgeID/GeolAgeDS). These links are compound (foreign key is based on two fields), with a '...DS' value that specifies the data container (such as a table or view in a relational database) for related data, and the '...ID' value identifies the particular linked data instance in that container. This convention holds for all table attributes that occur as pairs with the same name followed by the suffixes '...ID' and '...DS'.

The GeoUnitPartDescription entity correlates a geologic unit description with its parts, which may come from GeoUnitLexicon, GeoUnitDescription, NGMDB-CompoundMaterialLexicon, or CompoundMaterialDescription. Each part has an associated proportion, specified by a typical, minimum, and maximum value. The MeasuredQuantityTypeID attribute is a link to a science language term that specifies how the proportion values are to be interpreted—either as a value with uncertainty bounded by minimum and maximum values, or as a range with upper and lower bounds and a default typical value. Because the parts of a geologic unit may include other geologic units, the data structure is recursive. If the implementation constraints allow the most general application of this structure, standard SQL queries cannot be used to search for the occurrence of some particular lithology instance as a part of a geologic unit. The descriptions constructed thus far have avoided this difficulty by requiring geologic unit parts to come from a standard lithology lexicon.

A geologic unit also has an age, which may be specified in a number of ways, including a single term from a stratigraphic time scale (a geologic vocabulary), a range of ages specified by lower and upper bounds from a time scale, or a single numeric age (with error bounds) for an isotopically dated unit. In the implementation diagrammed in figure 6, geologic age is represented by a StratigraphicAgeRange entity that defines a lower and

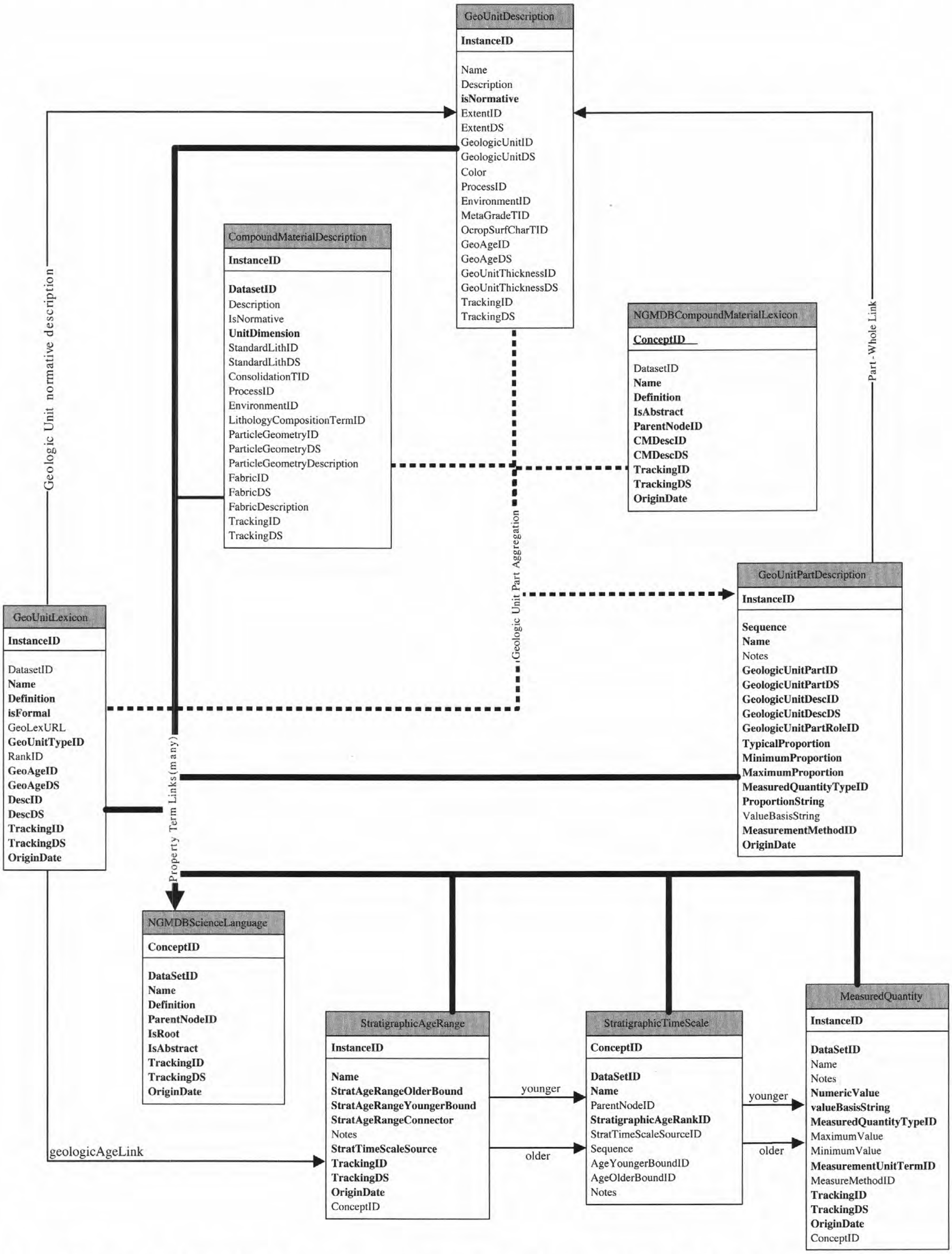


Figure 6. Entity-Relationship schema for geologic unit description. Pairs of fields with the same name with suffixes 'ID' and 'DS,' such as GeoAgeID/GeoAgeDS, are links to data that may come from different tables. The '...DS' value identifies the table that contains the linked record, and the '...ID' value identifies the appropriate record in that table. This mechanism allows for different representations of a property's value; for instance a geologic age may be a stratigraphic age range as shown in this diagram, but may also be reported as an isotopic age or a single stratigraphic age from a stratigraphic time scale. Other tables are used to describe geologic unit thickness, particle geometry, surface characteristics, and to record tracking (metadata) information. These have been omitted for simplicity.

upper age boundary specified using named ages from a stratigraphic time scale. The StratAgeRangeConnector attribute specifies the relationship between the age range bounds and the occurrence of a specified event during that time period (see table 6). The time scale is represented by the StratigraphicTimeScale table, which associates each named time interval (such as Jurassic) with lower and upper time boundaries in million years before present (Ma) specified by linked MeasuredQuantity records. The MeasuredQuantity table is a general representation for measurements, and allows recording of uncertainty bounds, range quantities, measurement method, and tracking information.

The link between geologic unit descriptions and the spatial data in a geodatabase is obtained using a relationship class that links the MapUnitID in the geology polygon feature class to the InstanceID in the GeologicUnitLexicon table. Other relationship classes in the geodatabase link the geologic unit to its geologic age, normative description, and rank definition. Because relationship classes do not allow joining on compound keys, the current implementation requires that all geologic unit definitions, geologic ages, and science vocabulary are each contained in individual tables. Some significant customization of the ArcGIS environment would be necessary to realize the full expressive power of the relational implementation outlined here. Because of the complexity of the knowledge representation necessary to fully capture geologic unit description and Earth material description in a computer analyzable form, other knowledge representation tools, including frame-based and description logic systems, are being investigated for data storage and analysis.

REFERENCES

NADMSC (North American Data Model Steering Committee), 2003, NADM conceptual model 1.0, A conceptual model for geologic map information: preliminary website release under the auspices of the Geological Survey of Canada, the U.S. Geological Survey, and the Association of American State Geologists. Available at <http://geology.usgs.gov/dm/steering/teams/design/NADM-C1.0/NADMC1_0.pdf>.

Richard, S.M. and Orr, T.R., 2001, Data structure for the Arizona Geological Survey geologic information system; basic geologic map data, *in* Soller, D. R., Ed., Digital Mapping Techniques 2001, Workshop Proceedings, U. S. Geological Survey Open-File Report 01-223, p. 167-188, <http://pubs.usgs.gov/of/2001/of01-223/richard2.html>.

Table 6. Age range relationship types.

Name	Definition
equal	Minimum and maximum ages are the same.
and	Unit includes rocks of both minimum age and maximum age, but not ages in between.
enum	Unit includes rocks of several discrete ages ranging from minimum age to maximum age; these are enumerated by AgeLabel.
or	Rocks in unit are either of minimum age or maximum age, or both ages.
through	Unit includes rocks continuously ranging in age from minimum age through maximum age.
to	Unit includes rocks discontinuously ranging in age from minimum age to maximum age.

APPENDIX

Data dictionary for tables in Figure 6.

Data types are as follows. Boolean, real number, and integer are standard data types. Short text is text string < 256 characters long. Long text is a text string longer than 255 characters, equivalent to an MS Access memo field. Links to a specified table (such as NGMDBScienceLanguage) are the datatype required by the unique identifier (key) for that table. The project is in the process of converting to system-assigned globally unique identifiers (GUIDs) as the key for data objects. Links to a table specified as a 'description object' are compound, with the first part identifying the record in the table (foreign key), and the second part identifying the table that is the source of the linked data. Appendix Table 1 lists fields that are present in all other tables.

Appendix Table 1. Fields that are present in all tables.

Attribute Name	Definition
ObjectGUID	Unique identifier for a data object in a dataset. Originally implemented as autonumber long integer, now converting to GUID (globally unique identifiers).
Name	Short text that specifies words used to identify a data object or observation instance. For many data objects, this represents a text summary of a complex value for display in reports or pick lists.
Tracking	Link to tracking record.
OriginDate	Automatic field, database supplies time stamp for creation of data object.

Appendix Table 2. CompoundMaterialDescription. Entity representing description schema for Compound EarthMaterial from NADM-C1 model.

Attribute Name	Definition
Description	Long text description, for use by human reader.
IsNormative	Boolean value, true if a description defines a term in some controlled vocabulary.
UnitDimension	Simple real number, >0, that defines the minimum size of sample necessary to characterize the material (give diameter of approximately spherical sample, in meters).
ConsolidationTerm	Link to term in NGMDBScienceLanguage that specifies the degree of consolidation of a compound material.
Color	Short text description of the color of some object.
Process	Link to term in NGMDBScienceLanguage that specifies a geologic process associated with the genesis of an EarthMaterial, GeologicUnit, or GeologicStructure, or that may be associated with a GeneticEvent in a Genesis description.
Environment	Link to term in NGMDBScienceLanguage that specifies a geologic environment (location) associated with the genesis of an EarthMaterial, GeologicUnit, or GeologicStructure, or that may be associatedwith a GeneticEvent in a Genesis description.
ParticleGeometry	Link to description object that specifies the geometry of all constituent particles in a compound material, or the geometry of some particular constituent particles in a compoundMaterial.
FabricTerm	Link to term in NGMDBScienceLanguage that specifies the fabric displayed by a compoundMaterial, or by one particular constituent in a compoundMaterial.
FabricDescription	Long text that provides description of fabric in a compound material or displayed by a particular constituent of a compound material.
StandardLithology	Link to most specific subsuming lithology term in NGMDBCompoundMaterialLexicon; use to simplify conflation of vocabularies or to relate lithology observation descriptions to a controlled vocabulary or lithology.
LithologyCompositionTerm	Link to term in NGMDBScienceLanguage to characterize general composition character of an EarthMaterial. More than one term may apply to a material. Includes rock names in chemical classification systems (for example, TAS).

Appendix Table 3. GeologicUnitDescription. Table summarizing major attributes of a bedrock or surficial geologic unit. Frequency on geologic unit part used to qualitatively express abundance, also allows to express constituents that are never present. Concrete table in NGMDB_P3 database only allows single values for Color, Process, Environment, and Metamorphic grade.

Attribute Name	Definition
Description	Long text description, for use by human reader.
IsNormative	Boolean value, true if a description defines a term in some controlled vocabulary.
GeologicUnit	Link to described geologic unit in GeologicUnitLexicon, if the description is associated with a named unit. If description is normative, link is to defined unit.
Extent	Link to an extent description object or directly to a spatial object that specifies the geographic region over which some description applies.
Color	Short text description of the color of some object.
Process	Link to term in NGMDBScienceLanguage that specifies a geologic process associated with the genesis of an EarthMaterial, GeologicUnit, or GeologicStructure, or that may be associated with a GeneticEvent in a Genesis description.
Environment	Link to term in NGMDBScienceLanguage that specifies a geologic environment (location) associated with the genesis of an EarthMaterial, GeologicUnit, or GeologicStructure, or that may be associated with a GeneticEvent in a Genesis description.
BodyGeometry	Link to term in NGMDBScienceLanguage that specifies the geometry of a geologic unit considered as a distinct body of material.
MetamorphicGrade	Link to term in NGMDBScienceLanguage that specifies the field metamorphic grade of a Rock or GeologicUnit.
GeologicAge	Link to geologic age description object that may be in StratigraphicAgeRange, GeochronDate (not depicted here), or StratigraphicAge table.
GeologicUnitSurfaceCharacter	Link to geologic unit surface character description table (not depicted here, includes attributes for surface dissection, varnish development, soil development, etc.).
GeologicUnitThickness	Link to geologic unit thickness description entity (not depicted here).

Appendix Table 4. GeologicUnitLexicon. Standard vocabulary of Geologic unit names that have an associated normative description.

Attribute Name	Definition
Definition	Long text used to specify the definition of some concept, for use by human readers.
IsFormal	Boolean value, true if geologic unit is formally defined.
GeolexURL	Short text that specifies a universal resource locator (URL) for the Geolex record associated with a geologic unit.
GeologicUnitType	Link to term in NGMDBScienceLanguage that defines the type of a geolgoic unit. Required term specifies type according to North American Stratigraphic Code. Other type terms may classify the unit according to other criteria (genetic process, environment.
Rank	Link to term in NGMDBScienceLanguage that defines the stratigraphic rank of a geologic unit.
GeologicAge	Link to geologic age description object that may specify a stratigraphic age, stratigraphic age range, isotopic date, or general age range. Schema shows implementation of stratigraphic age range with link to StratigraphicAgeRange table.
GeologicUnitNormativeDescription	Link to GeologicUnitDescription that is the normative description for the geologic unit.

Appendix Table 5. GeologicUnitPartDescription. Entity that contains properties of a part of a geologic unit. Tracking is inherited from GeologicUnitDescription. A GeologicUnitPart is always associated with a GeologicUnitDescription.

Attribute Name	Definition
Sequence	Integer to order a collection of data objects.
GeologicUnitPart	Compound link to geologic unit or earth material lexicon or description object; identifies geologic unit part. May link to GeologicUnitLexicon, GeologicUnitDescription, NGMDBCompoundMaterialLexicon, or CompoundMaterialDescription.
GeologicUnitPartRole	Link to term in NGMDBScienceLanguage, specifies role of a geologic unit part in the entire geologic unit.
GeologicUnitDescription	Link to GeologicUnitDescription from a geologic unit part; links part to whole.
TypicalProportion	Real number between 0 and 1, typical or default value to use for the proportion of a constituent in an aggregation.
MinimumProportion	Real number between 0 and 1, minimum value to use for the proportion of a constituent in an aggregation.
MaximumProportion	Real number between 0 and 1, maximum value to use for the proportion of a constituent in an aggregation.
MeasuredQuantityType	Link to term in NGMDBScienceLanguage specifying types of measured quantities—such as range, value with symmetric uncertainty, value with assymetric uncertainty.
ProportionString	Short text summarizes the proportion attribute for a constituent, for display in text controls.
ValueBasisString	String specifying how a default or typical value for a measured quantity was selected. Need to see what gets put in here to generate a controlled vocabulary.
MeasurementMethod	Link to term in NGMDBScienceLanguage that specifies the method used to determine a measured.

Appendix Table 6. GeologicUnitThicknessDescription. Description of thickness of a geologic unit--may be individual bed(s), or entire unit. Thickness quantity is multiple to allow stating things like 'normally this thick', never less than this thick'…

Attribute Name	Definition
Notes	Long text, for use by data compiler to enter general comments that pertain to a data object.
Extent	Compound link to an extent description object or directly to a spatial object that specifies the geographic region over which some description applies.
GeologicUnitDescription	Link to GeologicUnitDescription that thickness measure is associated with.
MeasuredQuantityType	Link to term in NGMDBScienceLanguage specifying types of measured quantities—range, value with symmetric uncertainty, value with assymetric uncertainty.
ThicknessSummary	Short text, original free text description of geologic unit thickness if from published source; summary text to display thickness value.
NumericValue	Numeric value for typical or default value of a measured quantity. Method of determining typical value specified by ValueBasis.
MaximumValue	Numeric value for maximum value in range, or upper bound on error envelope.
MinimumValue	Numeric value for minimum value range, or lower bound on error envelope.
ValueBasisString	String specifying how a default or typical value for a measured quantity was selected. Need to see what gets put in here to generate a controlled vocabulary.
MeasurementMethod	Link to term in NGMDBScienceLanguage that specifies the method used to determine a measured.
MeasurementUnitTerm	Link to term in NGMDBScienceLanguage that identifies the units of measure for an associated measured quantity.

Appendix Table 7. MeasuredQuantity. Entity for representing measured quantities that have associated units, measurement method, quantity, type. May be value with uncertainty; upper and lower bounds define uncertainty envelop, may be assymetric. May be range, with typical (default) value.

Attribute Name	Definition
Notes	Long text, for use by data compiler to enter general comments that pertain to a data object.
MeasuredQuantityType	Link to term in NGMDBScienceLanguage specifying types of measured quantities—range, value with symmetric uncertainty, value with assymetric uncertainty.
NumericValue	Numeric value for typical or default value of a measured quantity. Method of determining typical value specified by ValueBasis.
ValueBasisString	String specifying how a default or typical value for a measured quantity was selected. Need to see what gets put in here to generate a controlled vocabulary.
MaximumValue	Numeric value for maximum value in range, or upper bound on error envelope.
MinimumValue	Numeric value for minimum value range, or lower bound on error envelope.
MeasurementUnitTerm	Link to term in NGMDBScienceLanguage that identifies the units of measure for an associated measured quantity.
MeasurementMethod	Link to term in NGMDBScienceLanguage that specifies the method used to determine a measured.

Appendix Table 8. NGMDBCompoundMaterialLexicon. Controlled vocabulary of compound material definitions.

Attribute Name	Definition
Definition	Long text used to specify the definition of some concept, for use by human readers.
ParentTerm	Link to subsuming concept in this vocabulary.
isAbstract	Boolean, true if a term is 'abstract' (such as the name of a vocabulary) and can not be used to populate an attribute.

Appendix Table 9. NGMDBMineralLexicon. Controlled vocabulary of mineral definitions.

Attribute Name	Definition
ChemicalFormula	Short text, chemical formula defining a chemical composition.
ParentTerm	Link to subsuming concept in this vocabulary.
Level	Link to term in controlled vocabulary specifying the rank of a mineral name, from Micronex beta distribution <http://www.micronex.ca/ from http://www.georeferenceonline.com>.

Appendix Table 10. NGMDBScienceLanguage. Terminology for description of geologic features

Attribute Name	Definition
Definition	Long text used to specify the definition of some concept, for use by human readers.
isAbstract	Boolean, true if a term is 'abstract' (such as the name of a vocabulary) and cannot be used to populate an attribute.
ParentTerm	Link to subsuming concept in this vocabulary.
Vocabulary	Link to term in NGMDBScienceLanguage defining the vocabulary that contains this term.
TrackingNote	Short text notes on origin of a science language term, for simplified tracking; includes reference to publication and person, date of entry.

Appendix Table 11. StratigraphicAgeRange. Entity to represent a stratigraphic age that has a lower and upper bound expressed by a link to a named time interval in a Stratigraphic time scale.

Attribute Name	Definition
Notes	Long text, for use by data compiler to enter general comments that pertain to a data object.
StratAgeRangeOlderBound	Link to StratigraphicTimeScale for named time interval that is older bounding interval for a stratigraphic age range.
StratAgeRangeYoungerBound	Link to StratigraphicTimeScale for named time interval that is younger bounding interval for a stratigraphic age range.
StratAgeRangeConnector	Link to term in NGMDBScienceLanguage that specifies the distribution of rock unit age in time during a stratigraphic age range interval, for example, 'to', 'and', 'or', 'through'. 'enum' indicates that more than two ages are assigned.
StratTimeScaleSource	Link to NGMDBScienceLanguage definition that defines a stratigraphic time scale.

Appendix Table 12. StratigraphicTimeScale. Ordered, hierarchical partition of geologic time, with named time intervals. Parent-child links are partonomy. Sequence is for sorting intervals of similar rank, and sorting time scale.

Attribute Name	Definition
ParentTerm	Link to subsuming concept in this vocabulary.
Sequence	Integer to order time intervals.
AgeYoungerBound	Link to MeasuredQuantity table that specifies the younger bound in estimated absolute years, for a stratigraphic time interval.
AgeOlderBound	Link to MeasuredQuantity table that specifies the older bound in estimated absolute years, for a stratigraphic time interval.
StratigraphicAgeRank	Link to term from NGMDBScienceLanguage that classifies rank of a named time interval unit such as eon, period, stage…

Portable Software Tools for Managing and Referencing Taxonomies

By Jordan Hastings
U.S. Geological Survey and
University of California Santa Barbara
Santa Barbara, CA 93106
Telephone: (805) 893-3663
Fax: (805) 893-8617
e-mail: jordan@geog.ucsb.edu

ABSTRACT

Much of scientific enterprise is concerned with categorization of phenomena. Categorization schemes, also known as taxonomies, are of three types: flat, hierarchical (singly-nested), or hetrarchical (cross-nested, or correlated). Only flat schemes are easily represented in traditional database management systems and their associated user interface systems. Hierarchies and hetrarchies require additional key fields and/or tables to encode the nesting; these artifacts are obtuse both in the database and on forms. Described here is a portable spreadsheet application, Science Language Interface Module (SLIM), and a companion user interface widget, Tree-Box, that facilitate management of and reference to taxonomies in relational databases, making use of well-established graphical and textual notations for nesting. Limitations of these simple software tools also are discussed, along with suggestions for their future development.

1. INTRODUCTION

Phenomena in the real world demonstrate amazing variety. To deal with this variety, the human mind has evolved an exquisite capacity for categorization. We rapidly learn to distinguish parents from strangers, bread from bricks, sunsets from firestorms. As the categories become increasingly numerous and subtle, we repeatedly reapply the conceptual process to the categories themselves, thereby achieving hierarchical structures. The Linnaean system in biology and the geologic time scale are two well-known examples of such categorization schemes, or *taxonomies.*

Still, because of their variety, real-world phenomena may not fit clearly and cleanly into any fixed taxonomy. The duck-billed platypus, for example, begs to be included in multiple categories, as do many geological phenomena, such as fossil casts (both biologic and lithologic records) and tufa rock (arguably both igneous and sedimentary material). In reality, such *hetrarchical* phenomena are the exceptions that prove the rule of their strict hierarchical brethren, which can be neatly categorized.

Curiously, the computer systems that have gained wide acceptance over the past few decades—particularly the relational database management systems—do not directly support taxonomies. The disparity between messy real-world phenomena and flat, neat database tables is, in fact, extreme. With programming, this disparity can be masked somewhat, but the computational gestalt is at odds with much of the scientific enterprise.

This article is structured as follows: Section 2 presents a review of relational database technology and outlines how taxonomic organization can be implemented on top of it; Section 3 presents a textual representation of taxonomies that is convenient for both humans and databases; Section 4 describes a software package for managing taxonomies, implemented as a Microsoft™ Excel spreadsheet application; Section 5 describes a companion TreeBox "widget" that can be used to explore/select from taxonomies in user-written applications; and Section 6 offers some self-assessment and directions for future work

2. RELATIONAL DATABASE TECHNOLOGY AND TAXONOMIC REPRESENTATION

A relational database (RDB) is an organized and integrated collection of data maintained in the formal mathematical structure of relations (Codd, 1970). A *relation* is an unordered collection of *tuples*, each of which is a fixed-length list of co-occurring values for atomic *properties*, or attributes. Conventionally, relations are represented in tables: tuples correspond to rows, and property names (or more generally, property *domains*) correspond to columns. The actual attributes are recorded literally in the table cells. Tables are not relations, however, since they have fixed row and column orderings; also, cells may be left empty (to indicate missing or null values), which is intractable mathematically.

A relational database management system (RDBMS) is computer software that implements and supports one or more RDBs, commonly providing both programmatic access to them and an over-arching graphical user interface (GUI). RDBMSs concern themselves with many practical details of multi-user and multi-site access to RDBs, including concurrency, networking, security, etc. For an excellent overview as well as in-depth treatment of RDMBS topics, see Date, 2000.

It is difficult to represent even relatively simple real-world subject matter in a single RDB table; rather, a number of inter-related tables are used. The *relationships* between tables are entirely different from the relations represented within the tables themselves—a common point of confusion for database practitioners at all levels. The most common relationship between two tables, an *(equi)join,* is established by matching attribute values of selected columns, called *keys*, between them. The matching rows from the tables participating in such a join effectively create "super-rows" of attributes from both tables, usually showing the matched key only once. For each super-row in the output table there can be zero, one, or many matching rows in each of the input tables. For the "one-to-one" and "one-to-many" cases, the joined attributes can be appended to existing rows in one table or the other; however, the "many-to-many" case, which in practice is the most frequent, requires an intermediary third table to be represented (Date, 2000, p.76), adding substantial complexity.

Relational table structures are lionized as a means of modeling the real-world because, in theory, they permit individual facts (relations) to be stated only once—"the truth, the whole truth, and nothing but the truth, so help me Codd", as wags like to put it. However, the price for this pulverization of reality is high: much "relational work" is required to reassemble reality according to the matching keys between all the little tables of facts.

In truth, the RDB paradigm is reasonably well-suited to fiat realities, such as bank accounts and business transactions; it is a much poorer fit for real-world phenomena, which are characteristically dynamic, imprecise and subjective. Practically all of geoscience is this way! Nevertheless, RDBs are made to work, for better or worse, because other database approaches, such as object-orientation and markup languages, are still too experimental, expensive, or distant from existing desktop applications to be viable alternatives.

Fortunately, taxonomies address a fiat categorization of real-world phenomena rather than the phenomena themselves, so these are structures that can be nicely modelled in an RDB. Three kinds of taxonomy will be considered here:

- **Flat**: simple lists of phenomena, such as the palette of Munsell colors
- **Hierarchical**: Cleanly nested phenomena, such as stratigraphic rock-rank or the geological time-scale (American Association of Petroleum Geologists, 1983)
- **Hetrarchical**: Cross-nested, or correlated, phenomena, such as rock units with multiple ranks, or mixed clastic/carbonate sediments.

Evidently, a flat list is directly realized in a flat table. A hierarchy, with its restriction to single nesting, can be represented by a self-related table, in which each entry except the top-most "links up" to its parent as shown in figure 1a. A hetrarchy requires two tables, one to record the basic terminology (nodes), as for flat lists and hierarchies, and the other to explicate their inter-relationships (links). Among several representations for hetrarchy, Brodaric and others (2002) make a cogent argument for the most verbose option, where each node is explicitly linked to all of its ancestors right up to the top(s) of the structure; their example is reproduced as figure 1b. This "ancestor tree" notation also facilitates generalization to higher taxonomic levels, for both hierarchies and hetrarchies.

3. TEXTUAL REPRESENTATION OF TAXONOMIES

The RDB representation of taxonomies is serviceable in a computer application but hardly succinct for everyday use, especially publication, since its interlinking key structures are obtuse. Instead, for many centuries, humans have used "dotted number" *tags* (sometimes dotted letters or even dotted Roman numerals) to indicate hierarchical structures, especially outline structure in documents. The outline of this article, for example, is flat: 7 numbered Section headings. The North American Stratigraphic Code (American Association of Petroleum Geologists, 1983) by contrast, has a 6-level structure, in which only some levels are tagged (bolded):

PART **II**: ARTICLES
    FORMAL UNITS DISTINGUISHED BY CONTENT, PROPERTIES, OR PHYSICAL LIMITS
        LITHOSTRATIGRAPHIC UNITS
            Nature and Boundaries
                Article **22**. Nature of Lithostratigraphic Units
                    a. Basic units

Following on from Johnson and others (1999), Johnson (2002) has specified a dotted-number tagged hierarchy for classifying rock units on geological maps, part of which is shown in table 1 and used in subsequent examples. The tags prefixing the term names and descriptions are arbitrary.

Table 1. Portion of a lithologic hierarchy

Tag	Term	Description
1.	Unconsolidated	A sediment that is loosely arranged or unstratified …
2.	Sedimentary rock	A rock resulting from the consolidation of loose …
2.1.	Clastic rock	A rock composed principally of broken pieces ...
2.1.1.	Mudstone	A general term that includes claystone, siltstone, …
2.1.2.	Fine-grained mixed clastic	A mixture of clastic sedimentary rocks varying …
2.1.3.	Sandstone	A medium-grained clastic sedimentary rock …
2.1.4.	Medium-grained mixed clastic	A mixture of clastic sedimentary rocks varying …
2.1.5.	Conglomerate	A coarse-grained clastic sedimentary rock …
2.2.	Carbonate	A sedimentary rock composed of more than 50% ...
2.2.1.	Limestone	A sedimentary rock consisting chiefly of calcite
2.2.2.	Dolostone	A sedimentary rock consisting chiefly of dolomite
etc.		

The textual representation using dotted-number tags is compact and convenient both for computer work and for human communications; it also serves well as an interchange (import/export) format between computers and/or database systems. A tagged taxonomy file contains lines of text that are identical to the rows of table 1, prefixed by the name of the taxonomy. Such files can be easily loaded into or dumped from the RDB formats discussed above.

To accommodate hetrarchy, I extend this notation to equated tags, or simply *equates*. Where a term has two or more tags, it occurs at multiple locations in a hierarchy—therefore a hetrarchy—simultaneously. Equates are given only with derivative occurrences of a term; the principal occurrence exists stand-alone. For example, the expression "2.1 = 1.2.3 Tufa: A chemical sedimentary deposit from geothermal water …" indicates that tufa occurs derivatively in the 2… hierarchy (perhaps as an igneous material) and also primarily in the 1… hierarchy (as a chemical sedimentary material). Multiply derivative equates also are permitted, e.g. "2.3 = 2.1, 2.2 Mixed Clastic/Carbonate …" Equated entries within a taxonomy may occur at parallel or different levels and may contain the same or different text (term names and/or definitions); i.e., equates strictly address structure, not content. As a complete example, one rendition of the Brodaric and others (2002) hetrarchy is shown in figure 1c.

4. MAINTAINING TAXONOMIES

The tagged taxonomy notation is easily managed in word-processing and spreadsheet software. Microsoft™ Excel is particularly useful for editing taxonomies since it interfaces well with RDBs such as MS Access (or any ODBC-compliant RDB, in fact) and a wide variety of other desktop applications. The Science Language Interface Module (SLIM) is just such an Excel application, written for the National Geologic Map Database Project (NGMDB) to facilitate its work with the many geoscience taxonomies that underlie geological maps. The SLIM software is available for demonstration and download from the NGMDB Web site <http://ncgmp.usgs.gov/ngmdbproject>, under the tools submenu.

On startup, SLIM displays its "cover sheet" (figure 2a), from which the application is controlled via the [Application] pad that appears at the right end of the Excel menu bar. Selections from this drop-down menu lead in general to sub-menus and/or dialog boxes, which are intended to be self-explanatory. For example, the Data|Access choice prompts to open a connection to an MS Access database that contains (or will be updated to contain) standard taxonomy and ancestor tree tables (per Section 2). Data|Export and Data|Import perform transport between a pre-connected database and tagged taxonomy text files (per Section 3). All menu choices are described in on-line help.

Once a database connection has been established, SLIM lists the names of the (usually multiple) taxonomies it contains in a drop-down box (figure 2a, lower center). Selecting one of these names causes the corresponding taxonomy to be generated on an additional worksheet, which also serves as its editing tableau (figure 2b). Tags appear in column A, any equates in column B, the defined term names in column C, and free-text descriptions (optional, but strongly encouraged) in column D. Edits to cells other than the tags are freely allowed; in addition, full rows may be cut-and-pasted within a taxonomy worksheet, or between worksheets, with automatic recalculation of tags (and equates, if necessary). These edits are immediately reflected in the internal cache and also updated to the connected database when it is closed.

Right-clicking in a cell may present a popup display, depending on the cell's column. From the Tag column,

Node ID	Parent ID	Term
1	0 (none)	Grandparent
2	1	Parent
3	2	Child
4	3	First Grandchild
5	3	Second Grandchild

a) Hierarchy, self-linked in a single database table (above)

Geologic Units (*root*)

- X_1 Tectonic Unit
  - X_2 Terrane
    - X_3 Formation
      - X_4 Member
- Y_1 Tectonic Unit
  - Y_2 Terrane
    - X_3 Formation
      - X_4 Member

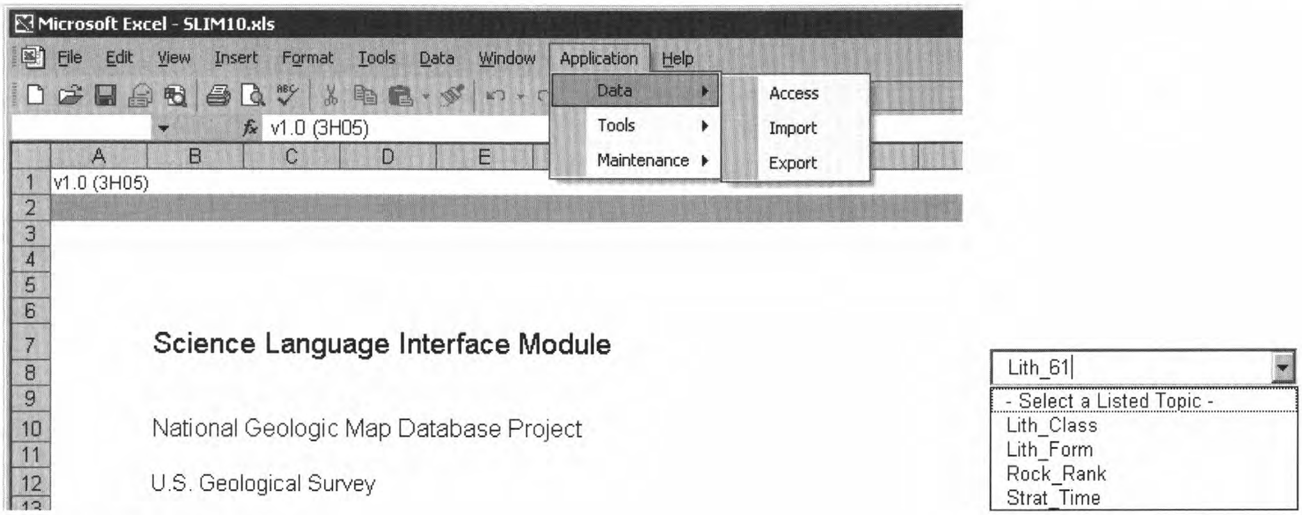
child	parent	edge
X_1	*root*	✓
X_2	X_1	✓
X_2	*root*	
X_3	X_2	✓
X_3	X_1	
X_3	*root*	
X_3	Y_2	✓
X_3	Y_1	
X_4	X_3	✓
X_4	X_2	
X_4	X_1	
X_4	*root*	
X_4	Y_2	
X_4	Y_1	
Y_1	*root*	✓
Y_2	Y_1	✓
Y_2	*root*	

b) Hetrarchy (X_3, X_4) as depicted explicitly (above), and expanded in database "ancestor" tree (right)

Tag	Term	Description
1.	X_1	Tectonic Unit
1.1	X_2	Terrane
1.1.1	X_3	Formation
1.1.1.1	X_4	Member
2.	Y_1	Tectonic Unit
2.1	Y_2	Terrane
2.1.1=1.1.1	X_3	*Formation*
2.1.1.1=1.1.1.1	X_4	*Member*

c) Same hetrarchy (X_3, X_4) shown as tagged list (left). Note that repeated terms are implied (italicized text)

Figure 1. Taxonomic representations.



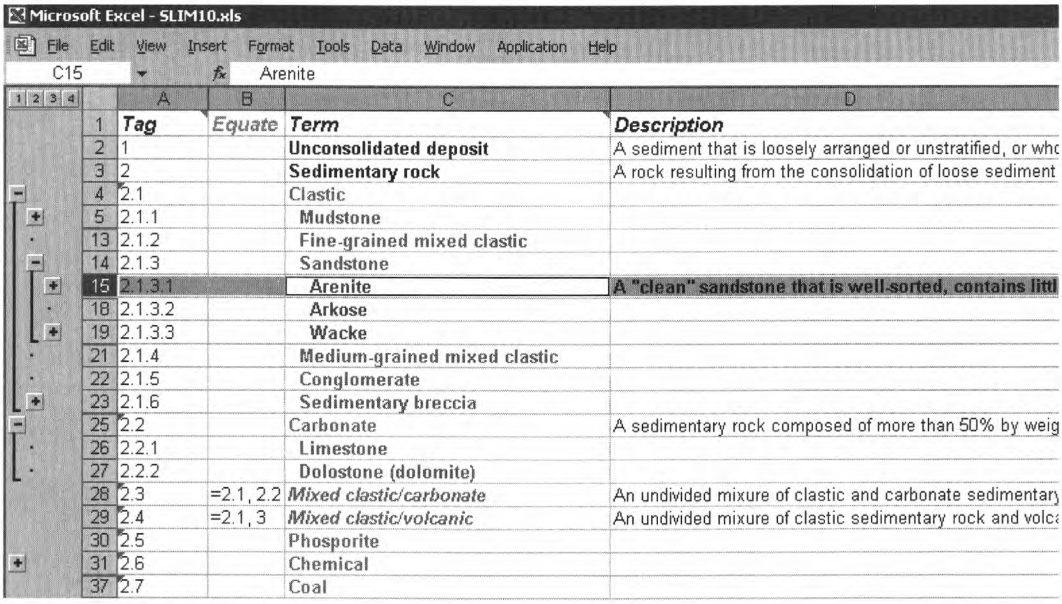
v1.0 (3H05)

Science Language Interface Module

National Geologic Map Database Project

U.S. Geological Survey

a) Application “cover sheet”



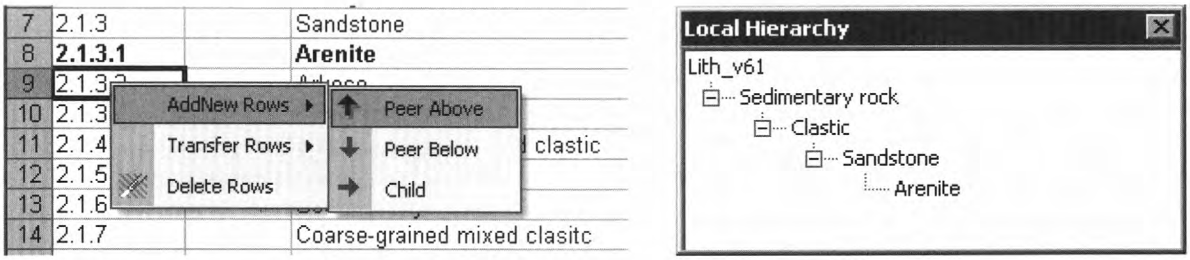
	Tag	*Equate*	**Term**	***Description***
2	1		**Unconsolidated deposit**	A sediment that is loosely arranged or unstratified, or wh
3	2		**Sedimentary rock**	A rock resulting from the consolidation of loose sediment
4	2.1		Clastic	
5	2.1.1		Mudstone	
13	2.1.2		Fine-grained mixed clastic	
14	2.1.3		Sandstone	
15	2.1.3.1		Arenite	**A "clean" sandstone that is well-sorted, contains littl**
18	2.1.3.2		Arkose	
19	2.1.3.3		Wacke	
21	2.1.4		Medium-grained mixed clastic	
22	2.1.5		Conglomerate	
23	2.1.6		Sedimentary breccia	
25	2.2		Carbonate	A sedimentary rock composed of more than 50% by weig
26	2.2.1		Limestone	
27	2.2.2		Dolostone (dolomite)	
28	2.3	=2.1, 2.2	*Mixed clastic/carbonate*	An undivided mixure of clastic and carbonate sedimentar
29	2.4	=2.1, 3	*Mixed clastic/volcanic*	An undivided mixure of clastic sedimentary rock and volc
30	2.5		Phosporite	
31	2.6		Chemical	
37	2.7		Coal	

b) Editing tableau

Figure 2. SLIM interface.

addition/deletion and transfer of terms is supported (figure 3a); and from the Term column, the local hierarchy associated with the term is shown (figure 3b). These same options are also available from the [Application] pad on the Tools submenu (not shown).

To facilitate editing in hetrarchies, several conventions apply. Derivative rows are shown italicized (figure 2b, bottom) vs. roman font for principal rows (as for flat and hierarchical taxonomies, all rows of which are principal). Rows may be converted from principal to derivative, and vice versa, simply by adjusting their equates; however, a row must remain principal, and also cannot be deleted, so long as it has any derivatives. Finally, when a new derivative row is added, it automatically copies up the (first) principal row’s term name and description, which subsequently may be edited.



a) Row editing (Tag column)

b) Local hierarchy (Term column)

Figure 3. SLIM interface popups.

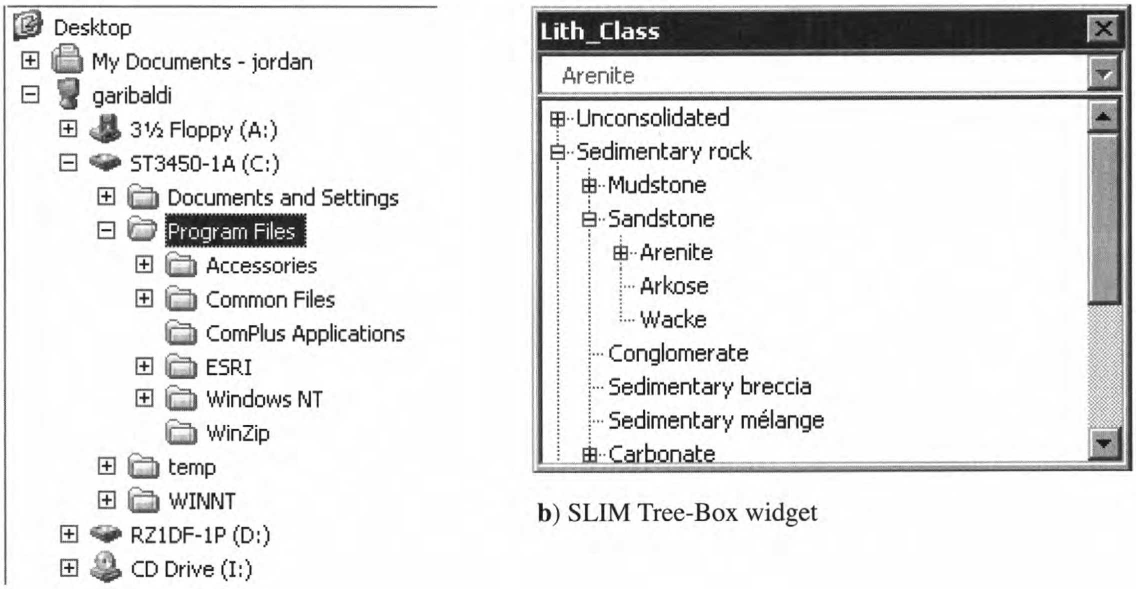
5. REFERENCING TAXONOMIES (THE TREE-BOX WIDGET)

Constructing taxonomies is, thankfully, only an occasional chore. Once constructed, good taxonomies can ease regular day-to-day tasks by providing hierarchical "pick-lists" that both speed up and standardize data entry and editing. For this purpose, SLIM includes a companion widget, called Tree-Box, which can be used to access its taxonomies from other applications. Only browsing and selection are allowed in Tree-Box, not editing.

The now well-established representation of hierarchies is a "tree-view", viz. the left-hand pane of Windows™ Explorer (figure 4a). In a tree-view, terms are appropriately indented, and [+] and [-] icons are provided for opening and closing, respectively, branches of the tree. Other established semantics include double-clicking terms in the tree to expand/collapse them, and right-clicking them to pop up auxiliary menus.

A miniaturized version of such a tree-view has been grafted onto a "combo-box" to make SLIM's Tree-Box widget (figure 4b). Tree-Box is implemented as an ActiveX control (Stephens, 1998) so that it can be easily embedded in MS Office™ applications or any similarly ActiveX-aware software, such as ESRI ArcGIS™. The relevant database connection can be set via Tree-Box's property pages at design-time and/or scripted at run-time. Subsequently, navigation of the selected taxonomy is under user control; any "picks" from it are returned to the host application as ordinary text strings.

Describing the behavior of Tree-Box is actually more complicated than using it. When closed, Tree-Box looks like an ordinary combo-box, which it is. Upon clicking the drop-down button, however, a tree-view pane is presented rather than a flat list; this can be navigated by the usual [+] and [-] icons, right-clicking, etc. Left-clicking on a node in the tree-view pops its name into the combo-box, and simultaneously closes the drop-down; to confirm



b) SLIM Tree-Box widget

a) Windows Explorer

Figure 4. Hierarchies as trees.

selection of the entry, a carriage-return is required. Alternatively, double-clicking in the tree-view pane, while open, selects and confirms a node selection in one step. Pressing ESCape at any time reneges the current operation, and Control-Z performs a full "undo" (single level).

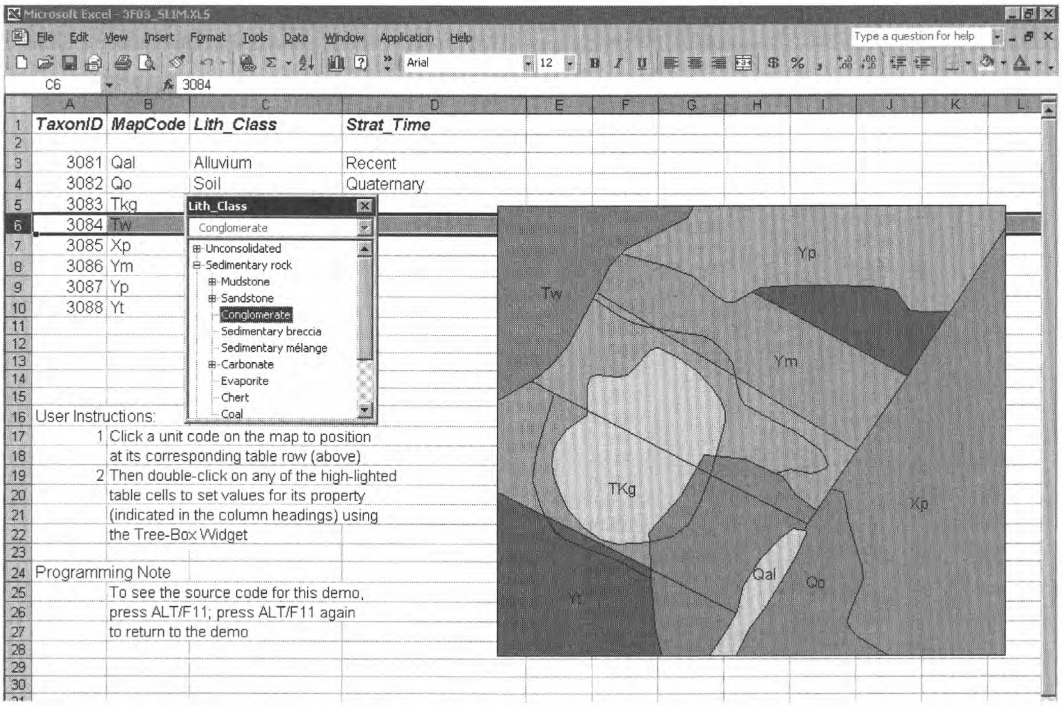
A demonstration of Tree-Box is incorporated into the SLIM application (figure 5); the stand-alone Tree-Box widget, as a self-registering .ocx file, also is available from the tools folder of the NGMDB project Website.

6. ASSESSMENT AND FUTURE WORK

Taxonomies and taxonomic thinking are endemic in the geosciences. It seems curious, therefore, that relatively simple tools like SLIM and Tree-Box are not more widely established. On the other hand, taxonomies (particularly in their hetrarchical form) are not straightforward to represent in a RDB; they also can become unwieldy in spreadsheet and word processing applications without database support. Taken together, the size and complexity of taxonomies can be daunting.

SLIM and its companion Tree-Box widget provide only a beginning point, not yet a compete system, for taxonomic processing. Important features not provided include:

- Auto-completion and searching: Terms typed directly into the Tree-Box are automatically compared against the underlying taxonomic list for matches. If an exact match is not made, however, it would be helpful for the drop-down panel to open with *all* the branches of the taxonomy that lead to possible matches already expanded, and with the partial matches themselves highlighted in tree-view. Achieving this functionality at reasonable computational cost is hard. Support for a more general search mechanism, perhaps utilizing *regular expressions*, could be helpful for large taxonomies.
- Change tracking and versioning: In organizational settings, it is customary to track changes to central databases, including those that support key infrastructure such as taxonomic lists. Tracking helps ensure that databases are up to date; that users have access to stable database versions, called "checkpoints"; and also that users can recover to such checkpoints in case of system failures. In distributed, evolving database environments, tracking multiple versions concurrently is often required.
- Cross-walking: Frequently, authors and/or organizations have investments in nearly equivalent taxonomies which they need to correlate or "cross-walk". Similar needs come about with multi-lingual taxonomies and often also with versioning (above). Support for manual cross-walks requires additional DB infrastructure—the correlations—which also must be tracked and versioned, etc. Automating cross-walks is a perennial topic of research in information retrieval.
- Extensions and personalization: People continually invent and nuance language. A fixed, rigid taxonomy can become boring as well as constraining to the scientific purpose. One concilia-



TaxonID	MapCode	Lith_Class	Strat_Time
3081	Qal	Alluvium	Recent
3082	Qo	Soil	Quaternary
3083	Tkg		
3084	Tw		
3085	Xp		
3086	Ym		
3087	Yp		
3088	Yt		

User Instructions:

1 Click a unit code on the map to position at its corresponding table row (above)

2 Then double-click on any of the high-lighted table cells to set values for its property (indicated in the column headings) using the Tree-Box Widget

Programming Note

To see the source code for this demo, press ALT/F11; press ALT/F11 again to return to the demo

Figure 5. Tree-Box demonstration, in SLIM.

> tory approach is a shallow taxonomy that permits users to add deeper terms for a particular project or purpose, or simply as knowledge expands. Allowing users to supply alternative or shorthand names for established terms, "nick-terms", which are supplanted by the proper terms on use, also can be a convenience.

Auto-completion and change tracking/versioning are essentially technical matters; cross-walking and dynamic extensions/personalization embed theoretical issues as well. The progress of science and language are clearly inter-related. Brodaric and Hastings (2002) argue that the relations are bidirectional, in fact. Specifically, the recognition and naming of new or different phenomena in the real-world eventually stimulates updated categorizations, for example, the recently-discovered sub-glacial lakes in Antarctica; while simultaneously, the naming conventions of established categorizations help focus attention on essential generalities and away from proprietary specifics, for example Munsell color (Munsell, 1946) vs. the plethora of other representations, RGB, HSV, CYMK, etc.

Finally, it is evident that the knowledge endeavor draws from many sources, many theories and many terminologies concurrently. A single taxonomy is rarely sufficient by itself, but tends to be inter-articulated and inter-constrained with other assertions of fact. For example, sedimentary grain size is a central axis of understanding for clastic rocks, but hardly relevant for chemical deposits; texture terms take precedence over composition terms for some types of rocks but not others. Encoding this kind of *peri-taxonomic* information is essential to the scientific process, even as the individual taxonomies are. Such ontologic considerations are beyond the scope of taxonomies *per se*, but are interwoven with them and through them.

In keeping with this self-assessment, SLIM and its Tree-Box widget have been made as simple as possible, initially, to learn from actual use what sophistications are needed/wanted. I welcome feedback concerning these prototype tools, especially suggestions for their improvement.

7. REFERENCES

American Association of Petroleum Geologists, 1983, North American stratigraphic code: American Association of Petroleum Geologists Bulletin, v. 67, no. 5, p. 841-875, <http://www.agiweb.org/nacsn/code2.html>

Brodaric, B., Boisvert, E., and Patera, A., 2002, A set-theoretic technique and software for managing multiple-classification in geological hierarchies, *in* Eighth Annual Conference of the International Association for Mathematical Geology, Berlin, Germany, September 15-20, 2002.

Brodaric, B., and Hastings, J., 2002, An object model for geologic map information, *in* Advances in Spatial Data Handling: Tenth International Symposium on Spatial Data Handling, Ottawa, Canada, July 9-12, 2002, p.55-68.

Codd, E.F., 1970, A relational model of data for large shared data banks: Communications of the Association for Computing Machinery, v. 13, no. 6, p. 377-387.

Date, C.J., 2000, An introduction to database management systems (7th ed.): Reading, Massachusetts, Addison-Wesley, 938 p.

Johnson, B.R., 2002. "Geologic map unit classification ver. 6.1", *At* <http://geology.usgs.gov/dm/>

Johnson, B.R., Brodaric, B., Raines, G.L., Hastings, J.T., and Wahl, R., 1999, Digital geologic map data model ver. 4.3, <http://geology.usgs.gov/dm/model/Model43a.pdf>

Munsell, A.H., 1946, A color notation (10th ed.): Baltimore, Maryland, Munsell Color Company, 74 p.

Stephens, R., 1998, Custom controls library: New York, New York, Wiley, 684 p.

CGMW Working Group on Standards for Digital Geological Data ("DIMAS")

By Kristine Asch

BGR - Federal Institute for Geosciences and Natural Resources
Bundesanstalt für Geowissenschaften und Rohstoffe
Stilleweg 2, D - 30655
Hannover. Germany
Telephone: ++49 +511 643 3324
Fax: ++09 +511 643 3782
e-mail: Kristine.Asch@bgr.de

The Working Group on Standards for Digital Geological Data (DIMAS) was instigated at the General Assembly of the Commission of the Geological Map of the World (CGMW) in Paris in 2002. The overall aim defined at that time was "to investigate and implement common CGMW standards for digital geological data structures and to review future mechanisms for digital data dissemination" (Resolution No. 65, in Commission of the Geological Map of the World, 2002).

In spring 2002, DIMAS met at the Federal Institute for Geosciences and Natural Resources (BGR), Hannover, Germany, for its kick-off meeting (Asch, 2003). The Terms of Reference (see table 1), the organisation of the working group, and its aims and objectives were then defined.

Table 1. DIMAS Terms of Reference

CGMW-DIMAS (Digital Map Standards Working Group) Terms of Reference
1. Propose, develop and promulgate geological and tectonic standards (including metadata standards) for CGMW maps, products and projects, and develop procedures for the management, maintenance, and upgrade of those standards.
2. Focus work on straightforward international, not national, map standards (CGMW map scales from 1: 1.5 M to 1:25 M) and address data models, data structures, data formats, dictionaries/terms/abbreviations. (Presentation elements, such as colours and layout, and stratigraphic classification are not a responsibility of DIMAS at this time).

The priority aims of the group are to:

1) establish the standard components of 1:5 million-scale maps
2) establish practicably applicable templates ("profiles") for small-scale geological maps
3) set up a system to define and manage metadata for all CGMW maps.

1) Establish the standard components of 1:5 million-scale maps

DIMAS initially will focus only on geological maps published by the CGMW. These standard components will include a draft scheme for the tectonic elements of those maps, a metamorphic scheme in cooperation with the IUGS Metamorphic Subcommission, a sedimentary scheme perhaps based on that developed by the Bundesanstalt für Geowissenschaften und Rohstoffe (BGR) for the International Geological Map of Europe ("IGME5000"), and a draft scheme for miscellaneous map elements. These draft standard components will be tested on a 1:5 million-scale CGMW map and GIS (for more information about CGMW, see http: //ccgm.free.fr/).

2) Establish practicably applicable templates ("profiles") of small-scale geological maps

Substantial discussion took place on 1:5 million-scale map "profiles" (standard schemas for igneous, tectonic, metamorphic and sedimentary rocks, plus miscellaneous features). It was agreed that

DIMAS should be a in a position to bring forward recommended profiles/schemas by the 2004 meeting of the International Geological Congress (IGC). In these profiles, DIMAS initially will focus on bedrock geology. It was agreed that the approach of the BGR IGME 5000 project to sedimentary, igneous and metamorphic rocks should be adopted and that the scheme for miscellaneous map elements (information about topography, bathymetry, georeferencing, etc.) was regarded as a possible template structure for the others.

3) Set up a system to define and manage metadata for all CGMW maps.

Significant progress had been made on this through the BGS work on a general ISO/TC211 19115-compliant metadata system (see http://www.isotc211.org/). It was agreed that once a more mature version of the BGS system was available, it should be adapted for CGMW use. DIMAS members Per Ryhaug (Norwegian Geological Survey, NGU) and Ian Jackson (British Geological Survey, BGS) will lead this activity. The aspiration is to have the system available and populated by the 2004 IGC meeting.

A DIMAS website has been set up as a means of communication for the group to exchange data and ideas (see www.geology.cz/host/dimas.htm).

Members of DIMAS come from a variety of geological surveys and organisations worldwide:

Kristine Asch(Chair), Federal Institute for Geosciences and Natural Resources, Germany
Manie Byrnard, Council for Geoscience, South Africa
Frank Brassil, Geoscience Australia, Australia
John Broome, Earth Science Sector, Canada
Ian Jackson, British Geological Survey, UK
Dominique Janjou, Bureau de recherches géologiques et minières, France
Manuel Pubellier, École Normale Supérieure, France
Robert Tomas, Czech Geological Survey, Czech Republic
Per Ryhaug, Norwegian Geological Survey, Norway
David Soller, United Stated Geological Survey, USA
Bruno Vrielync, Université Pierre et Marie Curie, France
Koji Wakita, Geological Survey of Japan, Japan
A member from China may join the working group soon.

Progress on DIMAS objectives will be presented to the CGMW General Assembly at the 32nd IGC in Florence in 2004.

REFERENCES

Commission of the Geological Map of the World, 2002, CGMW General Assembly, Paris, January 31 – February 2002, Resolutions: Paris, CGMW Bulletin 51, p.19-27 (http://ccgm.free.fr/index_gb.html)

Asch, K., 2003, CGMW Working Groups on Standards- DIMAS and GEOTERM: Paris, CGMW Bulletin 52, p. 25-27

Preservation of Geoscience Data and Collections

By Linda R. Musser

Fletcher L. Byrom Earth and Mineral Sciences Library
Pennsylvania State University
105 Deike Building
University Park PA 16802
Telephone: (814) 863-7073
Fax: (814) 865-1379
Email: Lrm4@psu.edu

DATA AND COLLECTIONS IN PERIL

Every day, geoscience data and collections are in peril of being lost– through deterioration, lack of space, loss of equipment to read the data, or lack or loss of documentation or metadata. The sheer volume of geoscience data and collections are daunting (table 1). A 1997 survey by the American Geological Institute (AGI) identified those data and collections available to be transferred to a repository if one were available (table 2). Unfortunately, few state geological surveys have space available to accept these materials.

In 2001, the National Research Council formed a committee to investigate this issue and recommend solutions. Specifically, the Committee on the Preservation of Geoscience Data and Collections was asked to:

- Develop a strategy for determining which geoscience, paleontological, petrophysical, and engineering data to preserve.
- Examine options for the long-term archiving of and provision of access to these data.
- Examine three to five access and repository case studies as examples of successes and failures.
- Distinguish the roles of public and private sectors in data preservation.

In addition to documenting the extent and nature of the problem, the committee identified factors that led to the loss of geoscience data and collections. These included:

- Loss or lack of space.
- Inadequate supporting information (metadata).
- Retirement or departure of staff without recording their knowledge.
- Deterioration of materials or metadata over time.
- Failure to migrate data to newer, more stable media or to operate with current technology/hardware.
- Lack of understanding by decision-makers of the value of data and collections.

Table 1. Minimum estimates of the volume of geoscience data and collections in the United States (National Research Council, 2002).

Collections	Units	Total
Core (ice)	Tubes	14,500
Core (rock/sediment)	Boxes	8,015,715
Cuttings	Boxes	10,402,000
Thin sections	Slides	647,000
Washed residues		180,000
Other well record		2,045,000
Fossils	Specimens	122,935,000
Minerals/rocks	Specimens	828,000
Data		
Seismic (2-d)	Line-miles	357,020,300
Seismic (3-d)	Square miles	249,849
Velocity surveys	Paper/digital	87,500
Well logs	Paper/film/ tape/digital	46,021,700
Scout tickets	Paper//film	21,960,350
Geochemical analyses	Paper	1,750,000

The committee recommended that priority for preservation should be placed on geoscience data and collections that are in danger of being lost (National Research Council, 2002). The highest priority should be given to those data and collections that are well documented and impossible or extremely difficult to replace. In addition to establishing priorities, the committee developed recommendations regarding storage, curation, cataloging and indexing, access, and discovery and outreach of geoscience data and collections.

The Committee's report reinforced the need for adequate space and funding to preserve geoscience data and collections through a combination of new space, support for existing repositories, and the creation of partnerships and consortia among repositories. Regarding curation,

Table 2. Volume of geoscience data and collections available to be transferred to a repository (National Research Council, 2002).

Data source	Volume available
Cores	10,000,000 feet
Cuttings	2,500,000 boxes
Thin sections	30,000 slides
Seismic data (paper, film and digital)	102,500,000 line-miles or films
Related data	25,000 velocity surveys
Well logs (paper, fiche and digital)	7,100,000 logs, cards, or tapes
Scout tickets	2,500,000 paper and fiche
Geochemical analyses	50,000 paper

cataloging, and indexing, the Committee emphasized the need for more support for these value-added activities and recommended that methods be developed within the scientific community to recognize outstanding contributions to curation, cataloging, and indexing of geoscience data and collections. Recommendations related to access, discovery, and outreach activities included more funding to make indexes available via the Internet and promoting recognition of the value of geoscience data and collections via citation.

Since publication of the committee's report there has been progress on several of the recommendations. A task force has been formed to promote the citation of geoscience data and collections by the geoscience community. The energy bill, currently before Congress, requests $30 million per year for 5 years for the USGS to distribute to state surveys for preservation efforts. Efforts are ongoing in the private sector and via the AGI Foundation to raise funds for preservation, and a group has been formed to act as a national advisory board on preservation of geoscience data and collections.

GOOD PRESERVATION PRACTICES

Hopefully, these initiatives will prove successful and more funding and space will become available to support preservation and archiving activities. In the meantime, it is important to take steps now to preserve geoscience data and collections. Be aware of existing guidelines, or staff in your organization who may assist you in your efforts.

For example, are there existing records management guidelines that cover your data? Have you consulted with the state archives staff regarding assistance they may be able to provide to your agency? Are there policies or guidelines developed by others that would be useful? The resources and guides developed by the International Council on Archives, the Archaeology Data Service , and the Council on Library and Information Resources offer good advice and examples. Some tips for handling digital collections include:

- **Document the data (create metadata).**
  The committee noted the lack of documentation as a persistent and widespread problem. If data lacks adequate documentation, it is of limited use. At a minimum, keep some documentation about the data on paper or simply written on the storage medium itself. For example, it is helpful to have something readable by the human eye that tells you what the data is, what format it is stored in, and what equipment or operating system is required to read it.
- **Transform data into a common format.**
  Don't risk the data by leaving it in a proprietary format unless absolutely necessary. There is a reason that text format is still widely used.
- **Store the data on long-term storage media.**
  Examples include gold CDs, DVD, and, yes, paper. Also, find a safe place, preferably offsite, to store back-up copies of data in case of catastrophe.
- **Establish a schedule for refreshing and migrating the data.**
  Long-term storage medium doesn't mean it lasts forever. Every storage medium needs to be periodically checked and refreshed. Technology changes may require migration to an alternate medium to accommodate loss of hardware. (Example – how many people today own turntables that can play a 78-speed LP?)
- **Assign responsibility for preservation.**
  Make certain it is part of someone's job to monitor preservation.
- **Make knowledgeable decisions.**
  It is important to recognize that not everything is worth saving. This is why records management guidelines are useful.

SUGGESTED REFERENCES

Archaeology Data Service, 2003, Guides to Good Practice. <http://ads.ahds.ac.uk/project/goodguides/g2gp.html>.

Council on Library and Information Resources, 2004, CLIR Reports. <http://www.clir.org/pubs/reports/reports.html>.

International Council on Archives, 1997, Guide for Managing Electronic Records from an Archival Perspective. <http://www.ica.org/biblio/guide_eng.html>.

National Research Council, Committee on the Preservation of Geoscience Data and Collections, 2002, Geoscience Data and Collections - National Resources in Peril: Washington, DC, National Academy Press, 107 p., <http://www.nap.edu/catalog/10348.html>

Integrating the Data Repository with the Publication Process

By Andrew Moore[1], Linda Guay[2], Ross Murray[1]

[1]Geological Survey of Canada, Natural Resources Canada
[2]ESSInfo, Natural Resources Canada601 Booth Street,
Ottawa, Ontario
Canada
K1A 0E8
Telephone: (613) 943-0776
Fax: (613) 947-9518
Email: amoore@nrcan.gc.ca

INTRODUCTION

In 1999, the Government of Canada made a commitment to become known "as the government most connected to its citizens, with Canadians able to access all government information and services on-line at the time and place of their choosing" (1999 Speech from the Throne). The resulting Government On-Line (GOL) initiative is the plan to deliver the Government's programs, services and information through the Internet. The Geological Survey of Canada (GSC) through its parent agency, Natural Resources Canada, is moving in a very significant way to support this mandate. As a result, the information technology / information management (IT/IM) and geoscience communities have felt the pressure to organize a Geoscience Data Repository (GDR) in which to store the wealth of the GSC's services, knowledge, information, and data in a fashion that makes it easily accessible by the public and, in particular, by decision makers. The GDR is a network of servers, applications, and databases that make up the corporate archive; this infrastructure requires standards that are easy to implement through the use of common practices and methods.

Beyond the basic infrastructure issues and requirements involved in defining a corporate IT/IM strategy, a greater level of coordination is required between management, scientists, and IT/IM staff. This coordination will facilitate the 'corporate' archiving and on line access to GSC's information themes. Geological mapping is one such theme in which a distributed database based on "version 5.x" of the North American Data Model (http://geology.usgs.gov/dm/) is being developed to archive not only the scientific interpretation and geospatial components, but also a normalized science language that describes the key attributes of the geologist's interpretations in the geologic map explanation, or legend. In order to facilitate the loading of information to the GDR, a process is being developed to integrate this information population into the GSC publication process. As with any science-based organization, the publication process is very important in ensuring a consistent and high quality product. By developing a process that efficiently manages the publication flow and integrates the information used to develop the publication, the GSC can meet the increasing demand for our knowledge and quality information while fulfilling our GOL mandate.

DRIVERS

The potential impact of geoscience information on both traditional and non-traditional decision makers is greatly increased when it is made available through the Internet. Even traditional publications such as maps, reports, and bulletins can take advantage of the wider distribution and easier access that the Internet provides. This is an obvious advantage to management, and there has been a push to develop these Internet portals. Unfortunately, what is less obvious is the amount of time, effort, and planning that is required to build the infrastructure to support these portals. Using traditional GSC science products such as a surficial geology map as an example and following it through the process to make it digitally accessible through the Internet, the obstacles that currently exist in getting this information to the corporate Geoscience Data Repository (GDR) are quickly identified. This is primarily the result of a traditional publication process thrust into the information management world. The knowledge and information that we want to make available through the Internet first has to go through the publication process. The publication process ensures quality information by providing several edit and peer review steps. This system works relatively well for traditional paper products; however, the observed primary information and data collected and compiled during the science project is now as much in demand as the published interpretation. The demands for high quality, consistent surficial geology information come from many applications such as groundwater protection, industrial mineral management, protected lands, basic research, mineral exploration, engineering, and environmental assessment. Prior to digital collection and digital map compilation, such information resided in field books or on aerial photographs; therefore, access to the primary information was extremely difficult. Currently, the primary information is archived on an ad-hoc basis, as the formal process to manage it is only now being developed. In the surficial geology example, we wish to develop a science language, or taxonomy, that can be used to represent the key attributes of the map legend. The legend will remain the domain of the scientist in order to reflect his interpretation of the glacial history and geological story, but the science language will provide a standardized way of querying this knowledge.

In developing tools to allow geologists to 'parse' (extract a consistent terminology from common attributes) their legends so that the spatial objects (points, lines, and polygons), original legend, metadata, symbology, and science language can be archived in the GDR, we realized that much of this same information is required by different working groups in the publication process. Divisional management and Earth Sciences Sector Information (ESS-Info) Publications require a permission to publish (P2P) form; critical reviewers must also be able to access much of this information to review the science; Editorial and Cartography sections, the Book Store, and Earth Science Information Centre (ESIC) draw upon, modify, and add certain components of the publication / information set. The preparation of the digital information to be accessed through our GDR and the preparation of the interpretation for publication are not integrated in a common process. If we can coordinate these two activities, we can not only make the effort more efficient but we can also improve the quality of the information and streamline the publication process.

CURRENT SITUATION

Each of these administrative groups in the GSC has adapted to information technology, but each has developed a solution that fulfils its individual needs. Unfortunately these solutions are not integrated and so the process of preparing and publishing a map requires a significant amount of interaction with several individuals and groups. For example, a P2P form must be completed, which includes author, title, project, etc. A divisional archive may be used to track publications, and that database requires author, title, project, etc. As the publication moves to critical review, a form is completed that includes author, title, project, etc. Once through internal review, an Open File number is requested. This information is entered into a database and, again, it includes author, title, project, and so forth. This continues through several steps of the publication process, which requires several different data fields of information beyond author, title, and project. The re-entry of identical information at various steps of this process has resulted in different titles in different databases for the same publication, legend conflicts between Editorial and Cartography that can delay the release of the publication, and other inconsistencies that render the process less efficient and accurate than it can be.

There is currently a project involving the GSC and ESS-Info division to develop a distributed network of corporate servers to archive, through the GDR, the GSC's geologic information in consistent data models; the Cartography section is a registered International Organization for Standardization (ISO) shop that requires standard forms and procedures and is using a database to manage its production flow; the Publications section employs a database system to track and manage GSC's publications; the Book Store uses a database to manage its stock and to keep track of GSC's clients; and ESIC, formerly the GSC Library, manages a database also used as a key source in Internet metadata access. Science divisions are developing the means to archive the published information in standard data models and taxonomies, such as in the surficial geology project, that will reside on the GDR. GSC Projects and ESS-Info are currently populating the GDR with surficial and bedrock maps from Cartography that are formatted into the designed data model (NADM 5.x). The GSC is to develop tools (such as the legend parsing tool) to link this digital map information to the science language. Publications, Editorial, Cartography (all part of ESS-Info) and the GSC science divisions are currently reviewing the publication process, with the goal of producing an updated publication process that will adapt to the IM infrastructure being developed. This will produce a process map or diagram that identifies what is required at each stage of the publication process and the person with whom the responsibilities lie. In effect, it will embed the infrastructure work (the same parsing tool as above) into a corporate process (publication).

TECHNOLOGY

The technology (software and hardware) and expertise that currently exist in ESS-Info and GSC can support the integration of thematic database standards development and corporate processes development, and we are now working to build this system and implement it in the GSC. Figure 1 shows how such a process can support the requirements of the publication process while integrating the demands and inputs of the corporate archiving work. Starting in the upper left of the figure and, again using a surficial geology map as an example, the author logs in to the Intranet system and indicates that he/she will be submitting an Open File (OF) map for publication. A P2P form (fig. 1) is then displayed and the required information is entered by the author. This information is submitted through the form to GDR, indicated here as the "The

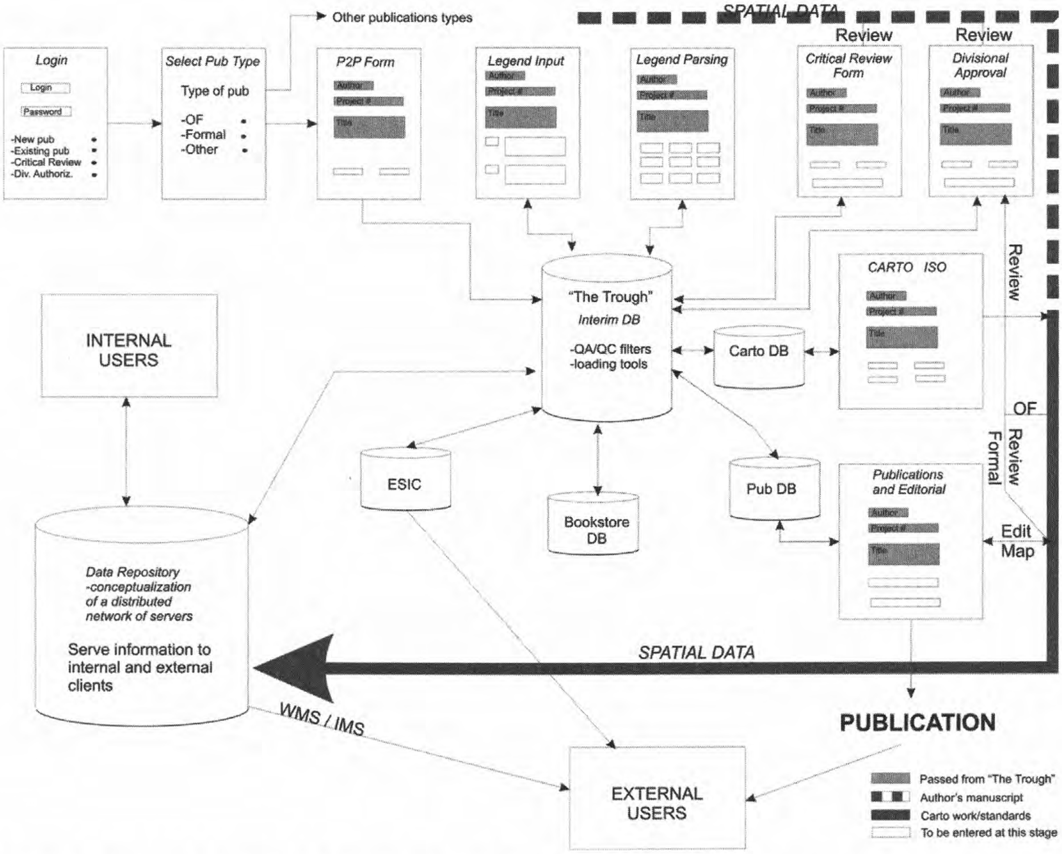


Figure 1. Planned publication flow, starting in the upper left part of the diagram and proceeding to the Geoscience Data Repository at lower left and publication at lower right.

Trough", for conceptual clarity and is simply a subset of the GDR that contains a variety of information needed for map preparation but flagged as unavailable for public consumption. In this process, the author ultimately uses the legend parsing tools to enter the legend description into the system and to parse the legend to the science language. As the publication and digital data move through the process, forms are displayed where many of the fields (such as the title) are already completed as the forms and/or local databases are being fed by the central database. The parsed information is also available for critical review in conjunction with the published product, providing a quality control step, and the database standard is then embedded in a corporate process.

Each step of the process triggers an action. For example, once the legend and P2Pform are submitted, an email is sent to division management indicating action is required on a publication. The sub-division chief reviews the P2P and assigns an internal critical reviewer. This triggers an email to the critical reviewer to review, and the process continues. Once Editorial and Cartography requirements are complete and divisional management has OK'd the publication (all electronically) the information is archived to the GDR, making it available to the public, and the publication is released. Each group or individual is assigned responsibility for each component of the information/publication set at each stage of the process.

REQUIREMENTS/RESULTS

The expertise to support this project is available in-house, but this project requires a significant amount of planning, cooperation, and effort. Each stakeholder (science divisions, Publications, Editorial, ESIC, and Cartography) must be consulted to ensure that each of their needs is met and that the responsibilities at each step of the process are clearly defined. For such a streamlined system to be effective, there must be 'buy in' from all groups involved. Consultation and participation by these groups is required in the planning and development of this system. Terrain Sciences Division has offered to pilot this process on behalf of the GSC science divisions. Implementing infrastructure standards of the IT/IM strategic plan will also support this project. For example, once the science language and data model for surficial geology are established, Terrain Sciences Division will not support a map publication until the legend has been parsed according to this corporate language. This will ensure that the publication process, and the archiving and access of our corporate knowledge are tightly linked.

A contract was awarded in December 2002 to build such a system based on commercial, off the shelf software (COTS) and custom modifications using standard tools and software. As of June2003, the initiative is in the final stages of user acceptance testing and will then be deployed as a pilot project using Terrain Sciences Division publications. Not only will the GSC be able to more efficiently produce publications and archive its vast information holdings, but the public will benefit by being able to access quality information and quality scientific interpretations in publications in a timely and coordinated fashion.

REFERENCES

Information Technology and Information Management Task Force, 2003, Earth Science Sector Information Management Plan: Earth Science Sector- Natural Resources Canada, internal document.

The Alaska Division of Geological & Geophysical Survey's Metadata Policy Development and Implementation

By Carrie L. Browne[1], Larry K. Freeman[1], and Gina R.C. Graham[2]

[1]Alaska Division of Geological & Geophysical Surveys
3354 College Road
Fairbanks, AK 99709-3707
Telephone: (907) 451-5031, (907) 451-5027
Fax: (907) 451-5050
e-mail: carrie_browne@dnr.state.ak.us; larry_freeman@dnr.state.ak.us

[2]Gina R.C. Graham
2575 Cresbard Ct.
Fairbanks, AK 99709
Telephone: (907) 474-3647
e-mail: graben@gci.net

ABSTRACT

In December 2001, the Alaska Division of Geological & Geophysical Surveys (DGGS) identified an emerging problem concerning the documentation of our geospatial datasets. For example, DGGS released 20 publications in 2001, of which 17 required metadata; however, only 1 actually included metadata. DGGS has published an enormous number of geospatial datasets in recent years, and the number of publications with geospatial datasets increases every year. Until recently there was no incentive, penalty, or process for documenting that data. To remedy this, DGGS launched a division-wide metadata policy. This policy stipulates that all future DGGS publications with geospatial data must have Federal Geographic Data Committee (FGDC) compliant metadata before they are published.

In January 2002, the DGGS director assembled a metadata committee to devise an efficient approach to generating metadata for our geospatial datasets. The committee proposed the establishment of an interim "transition period" of 6 months, January through June 2002, during which a minimum metadata standard would be met prior to publication. After June 2002, complete FGDC-compliant metadata would be required for all geospatial datasets prior to publication. Throughout the transition period, the metadata committee met semimonthly to familiarize themselves with the details of the FGDC Content Standard for Digital Geospatial Metadata Workbook version 2.0, and to discuss how the standard should be applied to DGGS. The committee's first priority was to determine which of the FGDC metadata elements specifically apply to the various types of geospatial datasets published by DGGS. Another task during the transition period was to research existing metadata generation tools and select the best tool for DGGS. The staff's diverse preferences in GIS software and geospatial data formats drove the metadata committee to recommend that DGGS create its own metadata tool specifically designed for the unique needs of its staff.

Metadata is now an integral part of the publication process at DGGS, as evidenced by the fact that, in 2002, 64 DGGS publications that required metadata did include it. We are continuing to improve the methods and have established an initial DGGS-specific metadata text template. We expect to have a user-friendly Web-based application available to DGGS personnel by October 2004.

INTRODUCTION

Since April 11, 1994, when President Clinton signed Executive Order 12906, the issues of data archiving and documentation have been brought to the attention of most local, state, or federal government agencies. One consequence of this Executive Order is that agencies must develop plans and allocate resources to generate metadata for geospatial datasets. If an organization introduces a metadata policy into their geospatial dataset creation process, managing the collection of their geologists' datasets becomes less cumbersome in terms of knowing what data they have, how it was generated, when, and by whom. One obvious scenario that illustrates this would be if a geologist were to leave a state survey after many field seasons. If the state survey required metadata documentation for all its geospatial datasets, it would be able to protect and manage the geologist's data left behind. The background information on how the dataset was generated would be preserved and the dataset would not be discredited by possible data misuse. Of course, the initial reaction of an organization's members learning they are now responsible for the generation of metadata might not be optimistic. Fortunately, this reaction is generally short lived because no one wants to have spent valuable time compiling a geospatial dataset only to have the data users discredit the dataset for lack of metadata documentation.

ESTABLISHING A METADATA POLICY

The Alaska Division of Geological & Geophysical Surveys (DGGS) has published geospatial datasets for more than 10 years and has, like most agencies, experienced challenges with its staff regarding the need to create metadata for all publishable geospatial datasets. Although DGGS staff knows why metadata is necessary and how to generate it for their geospatial datasets, very few older published geospatial datasets included metadata documentation. There were no real consequences to the staff for this lack of metadata, and so geospatial datasets commonly were published without adequate documentation. When a digital dataset was requested, metadata would have to be generated, sometimes years after the dataset was created and by individuals who didn't actually produce the data.

The DGGS metadata policy emerged in December 2001 after numerous public requests for digital geospatial datasets from DGGS publications that lacked metadata documentation. An investigation into our geospatial dataset documentation process was initiated to determine why none of the requested geospatial datasets included metadata. Our findings were staggering. During the 2001 publication year, only 1 of 17 publications containing geospatial datasets included proper metadata. Moreover, less than 5 percent of our published geospatial datasets over the previous 10 years included FGDC-compliant metadata.

We reviewed how DGGS had been addressing metadata creation and found that the agency was creating metadata for requested digital geospatial datasets on an *ad hoc* basis. Unfortunately, generating metadata this way was rapidly becoming overwhelming as the volume of geospatial data without metadata increased. We also realized that the originators of the geospatial data either had left the agency or were committed to other projects and were not available to assist in resurrecting information needed to write the metadata.

The lack of proper geospatial dataset documentation and the increasing pressure of public requests for our metadata were brought to the attention of the director of DGGS. In response, the director established a formal metadata policy stating that as of January 2002, all new geospatial datasets must include FGDC-compliant metadata before they will be published by DGGS. This formal policy was met with opposition and frustration at the survey. In addition, the initial metadata policy was flawed because it did not include an implementation plan. In response to the resistance to this policy, the director assembled a metadata committee from his staff to establish an effective work plan for generating FGDC-compliant metadata at DGGS. As a result of the staff discussions on metadata, it was clear that one goal of the committee would be to produce instructions and methodology for the generation of metadata that could be clearly communicated to project teams at DGGS.

METADATA TIMELINE FOR DGGS

The metadata committee consisted of DGGS managers and various staff members. The committee agreed that a strategy was needed to establish a transition into systematically generating metadata for our geospatial datasets. The recommended timeline was distributed at a survey-wide staff meeting in January 2002, and is presented in figure 1.

METADATA COMMITTEE MEETINGS

Once the survey staff agreed on the timeline, a second metadata committee was assembled. This committee was also staff-based and consisted of at least two members of each DGGS section (Minerals, Energy, Engineering Geology, and Geologic Communications). The committee met semimonthly for 6 months, during the timeline-defined "transition period," to become familiar with the details of the FGDC Content Standard for Digital Geospatial Metadata Workbook version 2.0 (Federal Geographic Data Committee, 2000) and to determine how to apply the FGDC standard to DGGS. Many of the committee members were not familiar with the FGDC workbook or its contents and had never completed a full FGDC-compliant metadata file for a geospatial dataset. To help get past our lack of metadata knowledge we used examples of metadata from USGS geologic mapping products, as well as some from other state surveys, and communicated by e-mail with Peter Schweitzer of USGS regarding metadata formatting and development questions that arose.

The first lesson for the committee members was to learn how to read the FGDC workbook and interpret its contents. Once everyone on the committee was familiar with the FGDC workbook we determined that our main concern was to select which FGDC metadata elements applied to specific types of geospatial datasets published by DGGS.

As the committee evaluated various metadata elements within the FGDC workbook, we began interpreting and clarifying relevant metadata element definitions to relate specifically to DGGS geospatial datasets. Our idea was that this strategy would assist data originators (DGGS

Timeline for DGGS Metadata Requirements

The following DGGS metadata requirements and timeline were discussed and agreed upon in a meeting on December 20. Present at the meeting were ADGGS director, section chiefs, and other interested members of the survey.

January 1, 2002 to June 30, 2002: Transition Period

- Minimum required fields for FGDC compliant metadata will be completed for geospatial data layers that were created for an ADGGS publication. This metadata will be considered preliminary and will include the fields listed on the "Interim ADGGS Minimum Required Metadata Fields" document (see attached template and example **).
- Preliminary metadata will be submitted to metadata coordinator for review to ensure that all interim minimums required metadata fields are properly documented.
- Publications submitted during this transition period will require a workplan to be submitted by the author to Geologic Communications section chief documenting the planned future date of completion of the final FGDC compliant metadata. The date of completion must be on or before October 31, 2002.
- Once the metadata has been reviewed by metadata coordinator, edited by publications editor, and the work plan documenting the future date of completion of the final FGDC compliant metadata has been approved by Geologic Communications section chief, the publication containing the data can be released, provided all other publication requirements are met.
- During this transition period, the Geologic Communications members with input from ADGGS staff will develop ADGGS-specific guidelines and templates to assist ADGGS staff in the creation of complete FGDC compliant metadata.

July 1, 2002: FGDC Compliant Metadata Required

- Beginning July 1, 2002, FGDC compliant metadata will be required for all geospatial data prior to its release in an ADGGS publication.

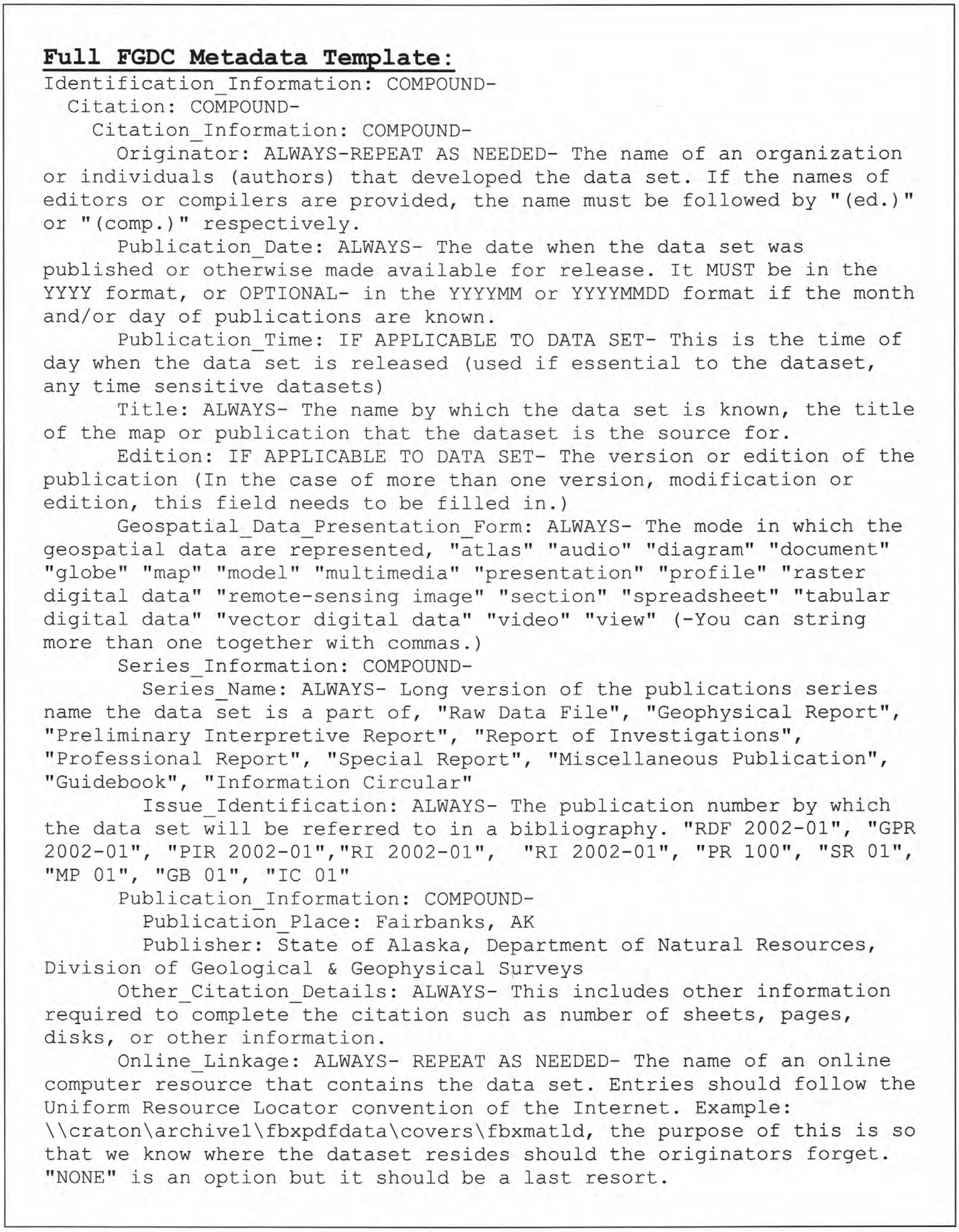
** The Interim template is very similar to Figure 2, see below
Written 1/3/02
Edited 1/4/02

Figure 1. DGGS timeline for implementing metadata policy.

geologists, staff, and other data contributors) in generating FGDC-compliant metadata for their unique geospatial datasets by helping determine when and why a particular element would apply to their dataset. The committee also supplemented metadata element domains with DGGS "boilerplate" text where appropriate or clarified element definitions so that data originators could clearly understand what was an appropriate entry for that element.

The committee thoroughly researched available metadata generation tools to determine the best tool for DGGS. We concluded that DGGS should create its own metadata tool designed specifically for its needs because the currently available metadata generation tools are software- and geospatial dataset-format specific, which does not allow for much flexibility. The metadata tool DGGS created is specialized to our diverse preferences in GIS software (ESRI ArcGIS, MapInfo, ERMapper, AutoCAD) and formats of geospatial datasets (geologic maps, data tables, databases).

The DGGS-specific metadata generation tool was produced in two phases. The first phase, completed in July 2002, consisted of constructing a text-formatted file or text template that includes all FGDC workbook elements and their DGGS-specific definitions applicable to DGGS datasets (fig. 2). In addition, DGGS provides training and support for those using this metadata template. The second, ongoing phase entails the production of a user-friendly interface application that is (1) not software specific, and (2), will output an FGDC-compliant metadata file when the data originator has finished entering information into the working metadata file. This tool is tentatively scheduled for completion by October 2004.



```
Full FGDC Metadata Template:
Identification_Information: COMPOUND-
  Citation: COMPOUND-
    Citation_Information: COMPOUND-
      Originator: ALWAYS-REPEAT AS NEEDED- The name of an organization
or individuals (authors) that developed the data set. If the names of
editors or compilers are provided, the name must be followed by "(ed.)"
or "(comp.)" respectively.
      Publication_Date: ALWAYS- The date when the data set was
published or otherwise made available for release. It MUST be in the
YYYY format, or OPTIONAL- in the YYYYMM or YYYYMMDD format if the month
and/or day of publications are known.
      Publication_Time: IF APPLICABLE TO DATA SET- This is the time of
day when the data set is released (used if essential to the dataset,
any time sensitive datasets)
      Title: ALWAYS- The name by which the data set is known, the title
of the map or publication that the dataset is the source for.
      Edition: IF APPLICABLE TO DATA SET- The version or edition of the
publication (In the case of more than one version, modification or
edition, this field needs to be filled in.)
      Geospatial_Data_Presentation_Form: ALWAYS- The mode in which the
geospatial data are represented, "atlas" "audio" "diagram" "document"
"globe" "map" "model" "multimedia" "presentation" "profile" "raster
digital data" "remote-sensing image" "section" "spreadsheet" "tabular
digital data" "vector digital data" "video" "view" (-You can string
more than one together with commas.)
      Series_Information: COMPOUND-
        Series_Name: ALWAYS- Long version of the publications series
name the data set is a part of, "Raw Data File", "Geophysical Report",
"Preliminary Interpretive Report", "Report of Investigations",
"Professional Report", "Special Report", "Miscellaneous Publication",
"Guidebook", "Information Circular"
        Issue_Identification: ALWAYS- The publication number by which
the data set will be referred to in a bibliography. "RDF 2002-01", "GPR
2002-01", "PIR 2002-01","RI 2002-01",  "RI 2002-01", "PR 100", "SR 01",
"MP 01", "GB 01", "IC 01"
      Publication_Information: COMPOUND-
        Publication_Place: Fairbanks, AK
        Publisher: State of Alaska, Department of Natural Resources,
Division of Geological & Geophysical Surveys
      Other_Citation_Details: ALWAYS- This includes other information
required to complete the citation such as number of sheets, pages,
disks, or other information.
      Online_Linkage: ALWAYS- REPEAT AS NEEDED- The name of an online
computer resource that contains the data set. Entries should follow the
Uniform Resource Locator convention of the Internet. Example:
\\craton\archive1\fbxpdfdata\covers\fbxmatld, the purpose of this is so
that we know where the dataset resides should the originators forget.
"NONE" is an option but it should be a last resort.
```

Figure 2. Metadata text-formatted template, including all FGDC sections/elements and their definitions, to be used when generating metadata for a geospatial dataset.

DEFINING A DATA UNIT

As the committee gained more knowledge about how to produce FGDC-compliant metadata, the question arose of how to divide a geospatial dataset to produce more helpful FGDC-compliant metadata. Establishing how many metadata files were needed for a large and diverse geospatial dataset was difficult. No one wanted to generate too many or too few metadata files for a geospatial dataset. A DGGS geologic publication consists of multiple layers or themes, each made up of a coverage or shapefile. If each coverage or layer of a published dataset were documented with an individual metadata file there would be numerous metadata files containing redundant data for each unique publication. On the other hand, if a single metadata file documents the dataset as a whole there is the legitimate concern that important pieces of metadata, essential to a specific layer or theme, may be lost. The metadata committee discussed three main options to address this problem.

Option 1: Publication Set = Data Unit

We proposed generating one metadata file for each published set of geospatial data and came to the following conclusions:

1) Only one metadata file would need to be generated for the dataset for this option.
2) The metadata file has the potential for being enormous and including a lot of information not applicable to any one thematic layer.
3) The types of links or relationships between elements of different sections of the FGDC standard necessary for this type of metadata file do not exist. This means a loss of data integrity between metadata sections/elements and specific thematic layers.
4) If a thematic layer was copied and reused from this publication set there is no way to parse the applicable metadata elements for that layer.

Option 2: Thematic Layer = Data Unit

Next we suggested generating one metadata file for each thematic layer composing a geospatial dataset. We concluded:

1) Each thematic layer would be fully documented by metadata with no loss of data integrity.
2) Many metadata files would have to be generated for one publication set.
3) There would be some redundant data contained in each metadata file within a subset of the same publication set.

Option 3: Distribution Determination = Data Unit

The last option that we considered was generating our metadata files as a function of how a digital geospatial dataset will be distributed. The following are our conclusions:

1) It would be easy for the originator of the dataset to determine how to break out the geospatial dataset into data unit(s).
2) Each geospatial dataset is fully documented by metadata with no loss of data integrity.
3) We may need to generate many metadata files that will include redundant data when a publication is distributed by thematic layer.

The metadata committee concluded that Option 3 was the best fit for DGGS. This means that the number of data units that make up the dataset (and thus, the number of metadata files required for a specific geospatial dataset) will be determined by how a digital geospatial dataset will be distributed by DGGS. For example, if a published geologic map will be distributed as a stand-alone map only, then the data unit would be the entire map and only one metadata file is needed for the publication. On the other hand, if a published geologic map will be distributed on a thematic layer basis when requested, then one data unit is assigned to each thematic layer.

METADATA GENERATION PROCESS

Once the text-formatted file including all FGDC workbook elements and definitions (fig. 2) became available for DGGS in July 2002, the survey began generating full FGDC-compliant metadata for our geospatial datasets. The metadata generation process is illustrated below in figures 3 and 4.

DGGS has an assigned Metadata Coordinator to check all metadata files for errors using the mp (metadata parser) tool (created by Peter Schweitzer of USGS). This saves time and effort by not trying to train everyone in how to interpret the error file mp produces. This mp program is a compiler to parse formal metadata, checking the syntax against the FGDC Content Standard for Digital Geospatial Metadata and generating output suitable for viewing with a web browser or text editor. It runs on UNIX systems and on PC's running Windows 95, 98, or NT. The mp tool generates a textual report indicating errors in the metadata, primarily in the structure but also in the values of some of the scalar elements (i.e. those whose values are restricted by the standard). We also have an onsite publications editor who thoroughly reads the metadata files for content and grammar to help eliminate errors.

Once the metadata file(s) have been generated for

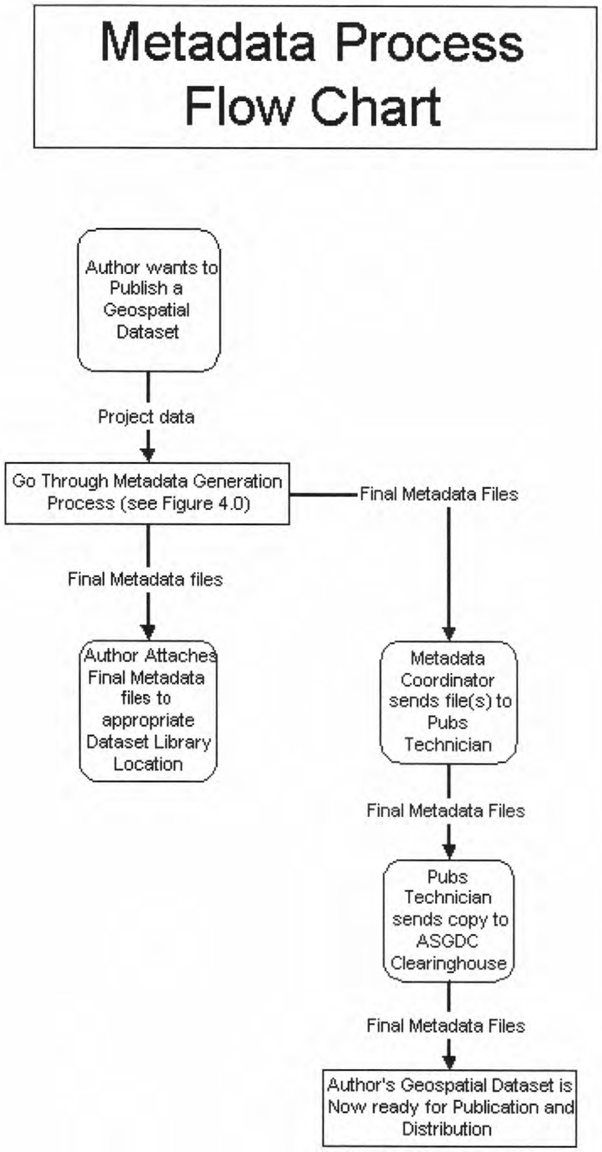


Figure 3. Metadata process flow chart providing an overview of DGGS metadata generation process.

a geospatial dataset they are sent to the Alaska State Geospatial Data Clearinghouse (ASGDC) to be posted on their website <http://www.asgdc.state.ak.us/>.

WHAT WE HAVE LEARNED

DGGS was not in a position in January 2002 to immediately begin generating proper FGDC-compliant metadata for its geospatial datasets. Even though the director made it a priority and policy that FGDC-compliant metadata must be included with all geospatial datasets published by DGGS beginning January 2002, the survey needed a strategy specifying how this new priority and policy would be implemented. This need is evidenced by the fact that metadata generation has occurred with greater frequency and precision since the metadata committee's strategy was introduced to the survey.

It is clear that a need exists for a user-friendly, non-software-specific application to help generate FGDC-compliant metadata. In all the research done by the second metadata committee, participants were astonished that with a federal metadata mandate in place there are few applications if any available that are non-software-specific and assist a user in generating metadata files.

The "FGDC Content Standard for Digital Geospatial Metadata Workbook version 2.0" is an invaluable resource for generating metadata for geospatial datasets. It is also a great resource for training personnel on what is expected of them when producing a metadata file. In DGGS, we decided to bring the workbook into a personalized language that the staff could understand and relate to in the form of a text-formatted metadata template.

The more documentation completed throughout the geospatial dataset production phase, the easier it is to generate metadata prior to the publication process. Once people understand what specific information metadata files require, they become more organized throughout the data gathering and documentation process of their project, resulting in more efficient metadata generation.

CONCLUSIONS

The DGGS metadata policy established in January 2002 sparked rapid development of FGDC-compliant documentation of all geospatial datasets ready for publication by DGGS. In the publication year 2002 DGGS generated 64 publications that contain or were produced from geospatial datasets; each is documented with metadata. The director-appointed division-wide metadata committee was successful in learning what is required to complete FGDC-compliant metadata. DGGS created a specialized metadata tool to aid in the completion of FGDC-compliant metadata and is currently producing a user-friendly, non-software-specific application that will be ready for use in October 2004. We have proposed a project and are applying for funding to generate metadata files for the legacy datasets from DGGS publications released prior to January 2002. If the project is funded, our entire geospatial dataset collection will include associated metadata files and be properly documented, archived, and protected for future use.

REFERENCES

Federal Geographic Data Committee, 2000, Content Standard for Digital Geospatial Metadata Workbook Version 2.0, Washington D.C., <http://www.fgdc.gov/metadata/meta_workbook.html>.

Schweitzer, Peter N., 1995, MP: A compiler for formal metadata (2.7.6 ed.): Reston, Virginia, U.S. Geological Survey, <http://geology.usgs.gov/tools/metadata/>.

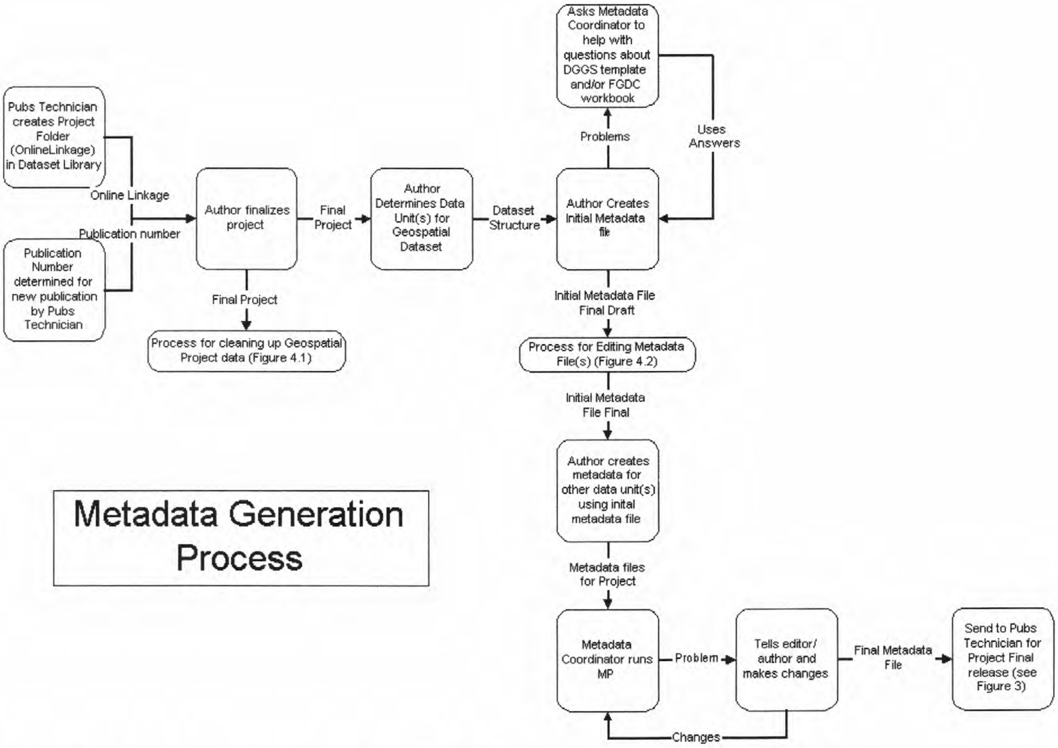


Figure 4.0. Metadata generation flow chart detailing all steps to be used when generating FGDC-compliant metadata for a geospatial dataset.

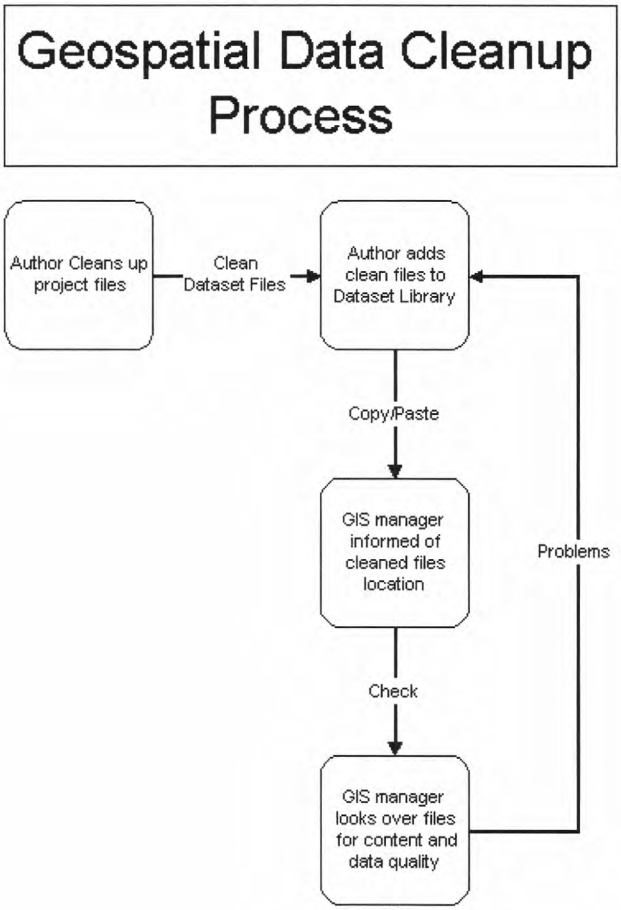


Figure 4.1. Geospatial data cleanup process detailing the checks and balances of our geospatial data files prior to storage in the dataset library.

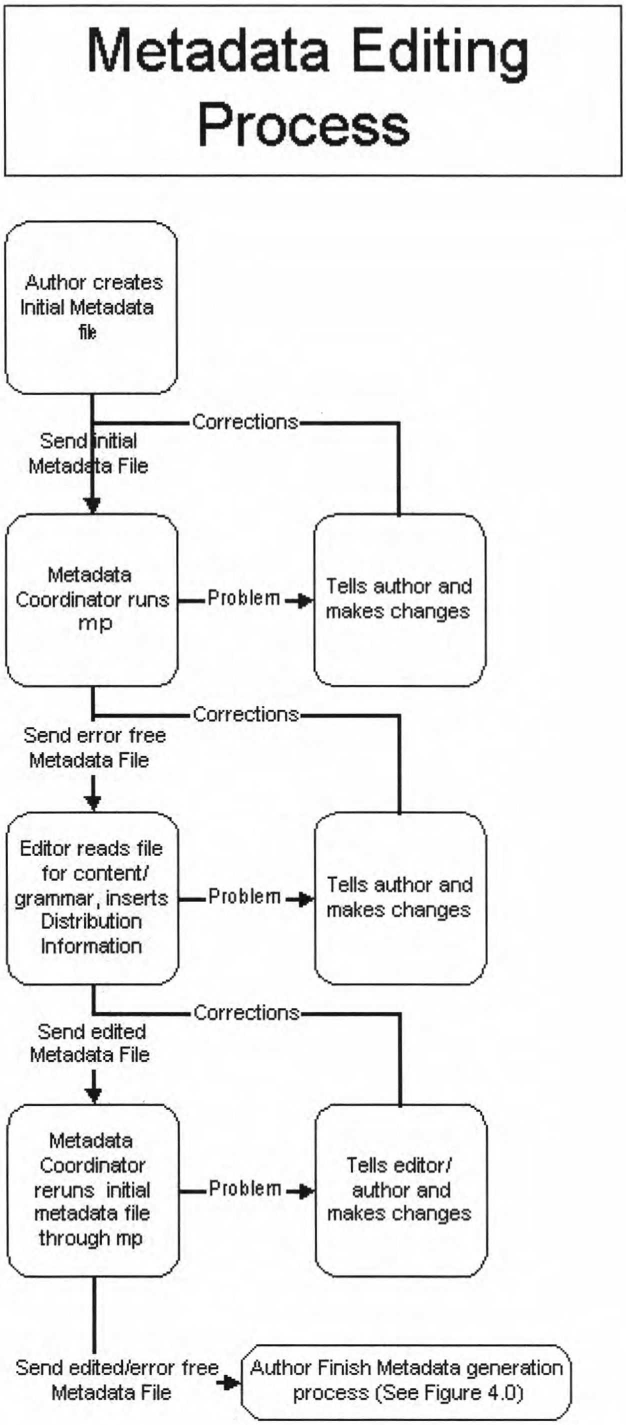


Figure 4.2. Metadata editing process detailing the steps a metadata file goes through at DGGS before it is considered FGDC-compliant metadata.

National Spatial Data Clearinghouse: The Real Story

By Peter N. Schweitzer

U.S. Geological Survey
Mail Stop 954
Reston VA 20192
Telephone: 703-648-6533
Fax: 703-648-6252
e-mail: pschweitzer@usgs.gov

INTRODUCTION

Since 1995, Federal and state agencies have put a lot of work into the National Spatial Data Clearinghouse. While that work has been fruitful, we've learned some useful things by looking closely at the character and usage of the clearinghouse and of the metadata it contains. The distributed search system designed for the national clearinghouse, though functional, is little used by the public it ostensibly serves. Are its contents therefore irrelevant? By no means! Usage statistics from a well-monitored clearinghouse site reveal that the public overwhelmingly prefers to use standard web search tools and local site navigation to find information. These observations support a new view of data catalog presentation that relies less on centralized search infrastructure, building more on the consistency inherent in metadata, increased use of controlled index terms, greater innovation in presenting information, and monitoring of actual use.

Geological surveys and other scientific organizations increasingly recognize the value of spatial data and the importance of well documented digital data for users both within and outside their walls. Efforts to communicate spatial data and metadata that cross organizational boundaries thus represent a common interest and a potential avenue for increasing the efficiency and improving the usefulness of the results of their research and monitoring. Once these efforts pass beyond the planning and promotional stages of development to operation, however, it becomes possible and important to evaluate their effectiveness in practice.

This report is intended to assess the effectiveness of the National Geospatial Data Clearinghouse search system from the point of view of an organization that distributes scientific data and metadata of broad appeal to the public.

BACKGROUND

With Executive Order 12906, signed by President Clinton in 1994, the National Spatial Data Infrastructure (NSDI) became a significant goal for Federal agencies and their funded cooperators, to which other organizations were encouraged to contribute. Briefly, the NSDI required data produced by Federal agencies to be documented using the metadata standard developed by the Federal Geographic Data Committee (FGDC), an interagency group charged with coordinating spatial data activities of, and cooperation among, Federal agencies. The metadata produced by the agencies was to be stored in a clearinghouse whose design was not finely specified in the order, and agencies were required to consult the clearinghouse before spending money to gather or purchase spatial data. The executive order also described the creation of the national geospatial data framework and stressed partnerships and standards development as well. This paper focuses on the spatial data clearinghouse.

Charged with the development of a clearinghouse that agencies would consult, yet not given the mandate to centralize spatial data delivery, the FGDC developed an architecture for the clearinghouse in which the metadata from different organizations would be stored and maintained in computer systems distributed widely across the country. Using internet protocols well established in the library computing community, the FGDC would devise search interfaces that would communicate requests from users to these widely-dispersed servers and, upon receiving results from them, pass the search results to the person who requested the information. This architecture is illustrated diagrammatically in figure 1.

The following factors that motivated FGDC to choose this clearinghouse architecture are important for the present discussion:

- People don't know where the servers are
- People don't know how to search each server if each server shows a different interface
- Why should everyone build a different search interface?

These concerns were and remain valid reasons to support, in principle, the distributed clearinghouse

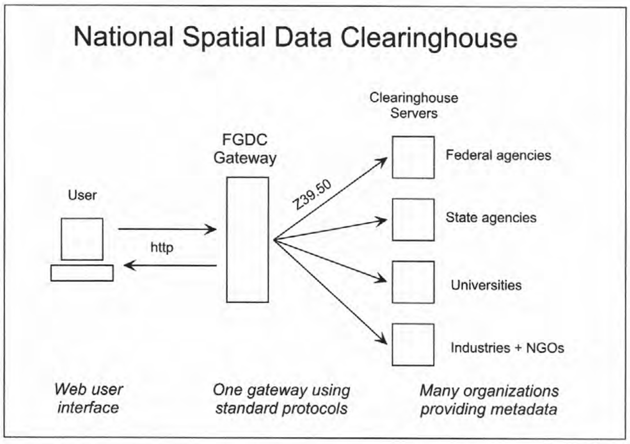


Figure 1. Diagram showing the architecture of the National Spatial Data Clearinghouse. The user wishing to search multiple metadata sources enters a dialog with the gateway machine, which communicates search requests to each of the clearinghouse's metadata servers according to the Z39.50 protocol, which is well-established in the library community. Responses are returned to the gateway and then to the user.

architecture with its centralized search interfaces on "gateway" systems.

CURRENT STATUS

Supporting the clearinghouse are many people engaged in several rather different activities. Writing metadata is perhaps the best-known of these, because it confronts scientific and technical experts with the unfamiliar and sometimes daunting structure of standard metadata. An organization with more than a few data producers will find that it needs to dedicate attention to the task of gathering metadata; arranging them in a collection; and imposing consistency in the expression of common terms such as publication series, keywords, disclaimer statements, format descriptions, and network addresses. The same people typically configure and run one or more clearinghouse servers, computers that understand the Z39.50 internet protocol used for searching distributed collections; and they often administer web servers distributing the same information using hypertext transfer protocol. Farther along the chain are those who design, set up, and run gateway systems, which provide the search interface that people can use to query the clearinghouse.

This report focuses on the concerns of the metadata collection manager and clearinghouse node administrator. In this role the chief concern is that the information to be provided to the public can be found, obtained, and used appropriately. Within USGS I carry out this function for geologic data. I begin this discussion with the frank admission that I find the search interface that FGDC has developed to be cumbersome and confusing. I believe that if alternative interfaces are provided to users they will be employed in preference to the clearinghouse interface.

TESTING THE EFFECTIVENESS OF THE CLEARINGHOUSE

For a person maintaining information to be made available to the public, the most important performance measure is to what extent that information is obtained by the public and the context within which the information is provided.

The USGS Geoscience Data Catalog is a collection of metadata records describing a wide variety of research results produced in the course of geological research conducted by USGS. These records are generally highly detailed and of excellent quality and consistency. At this writing the collection contains 1,589 records, but during the period in which statistics were gathered for this report the collection contained 1,117 records.

The records in this collection are accessible both using the Z39.50 protocol of the NSDI clearinghouse and using the typical hypertext transfer protocol (HTTP) used by most web sites. By examining the number and frequency of metadata records provided using these two protocols, we can learn how effectively they are distributing information to the public. An examination of this sort proceeds by analyzing in detail the log entries produced by the web server and Z39.50 server software.

LOG ANALYSIS

Specialized programs were written for these analyses because general log analysis programs confound{?} several unimportant influences in their results. For example, many requests for documents on the web are the result of automated "robots," also called "spiders," run by general search engines; the spiders read pages and the search engines create indexes of the contents of those pages to assess their relevance to queries that users submit. Similarly, many web sites log requests for files that are not complete documents but are ancillary information such as images. Likewise, web servers write log entries for requests even when the user already has a current copy of the information requested or when the information is known to be on a different server.

The web log analysis program written for this study examines all entries in the HTTP server log and counts only those requests that:

- *ask for bona fide metadata records*
  This is determined by examining the location of the requested file; on this server, metadata records are stored beneath the directory "metadata".
- *represent successful downloads*
  The HTTP server returns a result code "200" for

successful downloads; other codes are indicated for redirection or if the user already has an up-to-date copy of the file.

- *originate from users outside USGS*
  Since I am from USGS, this reduces the likelihood that the statistics are biased by my own download activity.
- *are not from a web spider or robot*
  This is accomplished by watching the user-agent field of the HTTP server logs. Every day a separate file is written containing only this field. By sorting it and eliminating duplicate entries, one can identify by inspection those user agents representing spiders, and store in a separate file a list of text strings that, if present in the user- agent field, indicate a spider.
- *result in complete downloads*
  The log records the number of bytes transferred; this is compared with the actual size of the file requested.
- *are not duplicate transfers*
  If the same host downloaded the same file on the same day, it is a duplicate and only one should count.

Source code for this program is available at:

http://geo-nsdi.er.usgs.gov/stats/yesterday.c
main program analyzes web logs
http://geo-nsdi.er.usgs.gov/stats/spiders.c
subroutines that read database of spiders
http://geo-nsdi.er.usgs.gov/stats/spiders.txt
database of web spiders

The programs are provided here for reference; they would need some customization in order to run on another computer system.

Metadata records on this server are available in several different formats: parseable (indented) text, outline-style HTML, Frequently Asked Question (FAQ)-style HTML, Standard Generalized Markup Language (SGML), and Directory Interchange Format. The output of the program `yesterday` shows the number of metadata records of each available format requested by real users outside USGS during the previous day. As ancillary information, the program also creates a file containing only the HTTP user-agent identifiers for these downloads, and another file containing the HTTP referrer, which is, for each download, the address of the web page containing the link the user clicked in order to download the metadata record. The summary of web downloads is updated daily and is available at http://geo-nsdi.er.usgs.gov/activity.shtml.

The Z39.50 server software used at this site is Isite, which was developed by the Center for Networked Information Discovery and Retrieval at the University of North Carolina and is maintained by the FGDC for use with the NSDI clearinghouse.(see http://clearinghouse4.fgdc.gov/ftp/) As distributed, the Z39.50 server software produces logs that are not sufficiently detailed to answer the question posed here; it records the search and present requests and the number of bytes transferred in the present request, but it does not record the type of record requested. Several record types can be requested: full, brief, and summary. Of these, only full records are significant. The brief and summary records provide only the document titles and are used by the gateway to compose a list of relevant documents which the user may request. To carry out this study, I modified the source code of Isite, recompiled it, and use the modified version to obtain more complete logs in Extensible Markup Language (XML) format. These modifications have been passed on to the maintainer of the server software, Archie Warnock, who has incorporated them into the current release. The modifications to Isite to improve logging are described at http://geo-nsdi.er.usgs.gov/stats/isite.html.

The modified Z39.50 server log is analyzed using a script written in PHP: Hypertext Preprocessor (PHP) (http://www.php.net/). The script creates an HTML document containing several tables:

- *Summary of activity*
  For each day, the number of search and present requests, the number of errors (mostly search requests specifying database fields for which we have no index or which are not present in the metadata), and the number of records presented, with brief and summary records tallied separately from full records.
- *Error messages*
  For the entire period spanned by the log, shows the number of times each error occurred. Few of these represent misconfigured requests; most are searches on fields that we do not have, such as the international standard book number (ISBN).
- *Clients generating present requests*
  For each remote computer that requests data, the number of requests, with the present requests broken down by type of record.
- *Clients generating search requests*
  For each remote computer carrying out a search, shows the number of searches and the number that could not be carried out due to errors.
- *Fields searched*

For the entire period spanned by the log, shows the number of search requests for each field. Some fields requested frequently are not contained in the database, so these requests cannot be filled. This script is run twice daily and the results are available on the web at http://geo-nsdi.er.usgs.gov/zlog.html

In order to keep the log to a manageable size and the analysis to a manageable time, the log file is changed each month so that the current log contains only activity from the current month; logs from previous months are stored separately.

RESULTS

To evaluate the relative effectiveness of the Z39.50 and HTTP mechanisms, we compare the number of full records downloaded through the Z39.50 server with the number of metadata records downloaded by real users through the HTTP server.

During the 142 day period from 7 January through 28 May 2003, 102,945 (full?) records were downloaded by real users through the HTTP server. In the same time period, 1,180 full records were downloaded through the Z39.50 server. Overall, 87 times more records were downloaded through HTTP than through Z39.50. (These were the statistics reported at the DMT meeting.) From 7 January through the end of August, 2003, there were 142,179 downloads by real users through the HTTP server and 1,589 downloads through the Z39.50 server, giving a ratio of 89:1 in favor of HTTP. More complete statistics through November 2003 are shown in the two tables below.

Table 1. Web access statistics, by month. Columns show the various file formats available; requestors may access a metadata record in any format. Outline and FAQ are styles of HTML. SGML is available but no hypertext links are provided to the files. DIF is the Directory Interchange Format (version 4) of the NASA Global Change Master Directory. Text refers to the simple format in which indentation is used to display hierarchical relationships among metadata elements. Most hypertext links on the web point to one of the HTML formats.

Date	Text	Outline	FAQ	SGML	DIF	Total
Jan-03	1,401	7,493	11,099	13	1,339	21,348
Feb-03	1,478	7,090	11,531	8	1,393	21,532
Mar-03	1,371	5,383	14,187	17	1,102	22,068
Apr-03	1,376	5,924	13,648	18	1,344	22,314
May-03	1,448	5,464	12,339	28	1,362	20,655
Jun-03	1,047	3,667	8,898	22	937	14,575
Jul-03	762	3,027	7,058	23	822	11,696
Aug-03	753	3,030	6,721	14	736	11,257
Sep-03	945	5,076	10,886	27	1,042	17,979
Oct-03	1,209	5,996	13,501	17	1,174	21,902
Nov-03	1,041	5,304	11,706	13	1,119	19,184
Cumulative	12,831	57,454	121,574	200	1,2370	204,510

Table 2. Cumulative statistics for Z39.50 access through November 2003. When a brief record is requested, only the document's title is provided; only requests for full records actually represent interest by a user. Errors are not necessarily improperly phrased requests, but include searches on fields that are not present in the database or which are not indexed.

Date	Requests		Error	Records presented	
	Search	Present		Brief	Full
Jan-03	13,634	705	6,599	6,340	285
Feb-03	14,819	917	6,463	6,149	311
Mar-03	17,576	1,381	7,323	5,178	175
Apr-03	17,292	1,042	7,206	4,842	219
May-03	15,163	826	6,068	4,319	202
Jun-03	14,218	426	6,456	3,082	118
Jul-03	12,318	771	4,586	2,925	108
Aug-03	10,969	395	4,210	3,275	170
Sep-03	13,293	675	5,890	6,410	160
Oct-03	14,109	430	5,271	4,433	83
Nov-03	12,468	413	5,662	3,539	146
Cumulative	155,859	7,981	65,734	50,492	1,977

The HTTP referer statistics for real users provide additional important information: The most frequent referer by a wide margin is the commercial search service "Google", but the second most frequent referer is a browse interface that is local to this data server: http://geo-nsdi.er.usgs.gov/cgi-bin/place

This browse interface allows people to locate metadata records by choosing from a limited set of commonly-known place names that are arranged hierarchically (continents, countries, states or provinces, and counties). This interface could be built only by ensuring that the metadata records used a rigidly consistent set of place keywords. Note that the FGDC metadata standard allows a record to be categorized using terms from many different controlled vocabulary as well as using non-controlled terms; the recommendation that consistent place keywords be used does not prevent other vocabularies—even other geographic terms—to be used as well, but the vocabulary that is used should be identified in the Place_Keyword_Thesaurus field.

DISCUSSION

In the development of the clearinghouse architecture, much attention was paid to the need for users to carry out a single search on numerous distributed servers, and to restrict the search to specific fields of interest. These are reasonable concerns that general Web-search engines cannot be expected to address. Notwithstanding these concerns, however, real people have chosen to use the web in preference to the clearinghouse by a wide margin. This finding implies that clearinghouse administrators who wish to maximize the distribution of their information to real users will:

- make their metadata available through the web,
- allow them to be indexed by common search engines, and
- build local browse and search interfaces.

From this analysis it is reasonable to ask whether continuation of Z39.50 service is cost-effective. The answer depends on several factors, the most important of which will vary from site to site; that is, the cost of administering the Z39.50 server software. In my experience this is not difficult, so I would not recommend that people who are already running the software discontinue it. But from the perspective of maximizing effectiveness, it is clear that the Z39.50 service is not contributing significantly to meeting the needs of the user community.

It should be noted that the large number of metadata records downloaded through the web indicates that this information is desired by users. If the experiment were simply looking at the number of downloads by Z39.50 on a single server, one might infer from a low number of downloads that people simply don't want this type of information. But since the same information is here available by a different method, that conclusion cannot be sustained. People want these metadata records, and they get them through typical web interactions, not through the clearinghouse.

The Z39.50 server is receiving a large number of search requests, yet is receiving few requests for full records. An examination of the search terms gives us some insight into this problem. Many searches appear to be requests for general topics, such as books and both classical and popular music. Indeed the most commonly searched field is the ISBN, or international standard book number; none of our data sets have this identifier. Many of these requests originate in university libraries, judging from the hostnames from which the searches originate. I believe that commercial software commonly sold to libraries is configured to query all available Z39.50 servers with all searches. It is therefore important not to regard the number or frequency of search requests as a measure of the effectiveness of the clearinghouse.

The clearinghouse search works, but people don't use it. In contrast, Web search wasn't expected to work effectively, but apparently it does. While it's tempting to blame the unpopularity of the clearinghouse interface on lack of publicity or the complexity of the search form, the explanation may be simply that people prefer the familiar to the unfamiliar, and that they will use such a system if it works well enough, even though a more complex, less familiar system would be arguably more comprehensive.

XML Encoding of the North American Data Model

By NADM Data Interchange Technical Team
Eric Boisvert[1], Bruce R. Johnson[2],
Peter N. Schweitzer[2], Martin Anctil[1]

[1]Geological Survey of Canada-Québec Division
Natural Resources Canada
880 Chemin Ste-Foy
Québec, Qc, G1S 2L2
Canada
Telephone : (418) 654-3705
Fax : (418) 654-2615
e-mail: {eboisver, manctil}@nrcan.gc.ca

[2]U.S. Geological Survey
National Center, MS 908
12201 Sunrise Valley Drive
Reston, VA 20192
Telephone: (703) 648-6051
Fax: (703) 648-6383
e-mail: {bjohnson, pschweitzer}@usgs.gov

INTRODUCTION

The North America Data Model (NADM) is composed of a series of initiatives to create a common set of tools and technologies to manipulate geological information in the digital realm. One of these initiatives is to define a standard interchange format to allow easy exchange of information between systems and tools.

Extensible Markup Language (XML) is a formalism to encode domain-specific information (such as chemistry, biology, recipes, geology, etc) into a structured document. XML encoding must follow certain rules to be both 'well formed' and 'valid'. Although creating a 'well formed' XML document is relatively simple, a 'valid' document must comply to the domain specific rules (for example., those for geology). The first draft of a conceptual data model for geology and geologic maps has been worked out by the North American Data Model Steering Committee's Data Model Design Team (NADM-DMDT)(2003). This conceptual model expresses the rules to which the science of geology adheres. This model will be translated into a set of XML document construction rules called an XML schema, which will be used to assess the validity of XML-encoded documents containing NADM-compliant datasets. The Data Interchange Technical Team (DITT) is working to develop this XML schema.

ACKNOWLEDGMENTS

Special thanks to Serge J. Paradis (Geological Survey of Canada) and David R. Soller (United States Geological Survey) for suggestions that greatly improved the manuscript.

DATA INTERCHANGE TECHNICAL TEAM (DITT)

The Data Interchange Technical Team (DITT) is one of the technical teams composing NADM. Their mandate is, among other things, to (see <http://geology.usgs.gov/dm/steering/teams/interchange/interchange.txt> for charter):

- Develop standardized formats and mechanisms for exchanging digital geologic map databases
- Facilitate exchange of digital geologic map content between various implementations of the NADM

Various technical solutions are available to achieve this task, but considering the current trend in information technology, every path seems to involve XML. In the fall of 2002, at the Geological Society of America (GSA) meeting in Denver, the DITT started working on the encoding of the conceptual model developed by the NADM DMT team. For this purpose, XML has quickly been elected as the technology of choice. This option was discussed early in the modelling process and the current technological trend only reinforced this option.

WHAT IS XML?

XML is a plain text file structured using 'tags', or '*mark-ups*', that organise the information contained in a document according to a set of predefined rules. Tags can be created to accommodate a domain and the rules to organise them can also be defined to fit the domain requirements.

XML technology has many advantages; it offers a general approach for structuring information in a document and is sophisticated and well adapted to the web environment. Lots of tools are available to manipulate XML and a growing community of XML enthusiasts, rooted in the *Open Source* movement, provides sufficient support (a search for XML on *Google* <http://www.google.com> returns 20 million pages !). XML is truly a multi-platform, multi-vendor, multi-programming language and even has its own transformation language (XSL / XSLT). Many emerging technologies are based on XML (for instance, the whole Microsoft Office XP suite) and it is at the root of the "Web Services" revolution.

Markup Language

A marked-up document is a file where important pieces of information are flagged to allow a human reader or a machine to quickly locate it, or to explicitly document it. For example, this small paragraph from Drewes (1998):

Dacitic Vent Breccia (Miocene)*—Light-medium-gray, finely porphyritic dacitic rock containing inclusions of Jurassic or Proterozoic granite and Jurassic rhyolite (welded tuff?) as much as 20 m in diameter. The subcircular outcrop mass of breccia probably is a volcanic vent or throat. A halo of strongly saussuritized rock 0.3–0.5 km wide*

makes sense for anyone who has a bit of geological background. One can easily locate the geological ages hidden in the text. A machine (or someone who has no formal knowledge in geology) cannot extract this information from the text because this piece of information is not explicitly identified. A software would need a complete thesaurus of age names. A simpler approach is to flag this information directly in the document:

```
Dacitic Vent Breccia (<age>Miocene</
age>)—Light-medium-gray, finely
porphyritic dacitic rock containing
inclusions of <age>Jurassic</age>
or <age>Proterozoic</age> granite
and <age>Jurassic</age> rhyolite
(welded tuff?) as much as 20 m in
diameter. The subcircular outcrop
mass of breccia probably is a volcanic
vent or throat. A halo of strongly
saussuritized rock 0.3—0.5 km wide
```

`<Age>` and `</Age>` are respectively opening and closing tags that flag subsets of the document and assign the enclosed words a specific interpretation; the tags identifies Miocene as an age. Better yet, attributes can be added to the tags to enhance the tag usability. For example, `lowerBound` and `upperBound` attributes contain the ages in millions of years before present, delimiting the stratigraphic age.

```
Dacitic Vent Breccia (<Age lowerBound="-
23.8" upperBound="-5.3">Miocene</Age>)—
Light-medium-gray, finely porphyritic
dacitic rock containing inclusions of
<Age lowerBound="-206" upperBound="-
144">Jurassic</Age> or <Age lowerBound="-
2500" upperBound="-543">Proterozoic</
Age> granite and <Age lowerBound="-206"
upperBound="-144">Jurassic</Age> rhyolite
(welded tuff?) as much as 20 m in diameter.
The subcircular outcrop mass of breccia
probably is a volcanic vent or throat. A
halo of strongly saussuritized rock 0.3—0.5
km wide
```

A second level of information structure is given by the organization of the tags themselves. In the previous example, the tags are scattered loosely in the text, but XML documents can be organised in a more strict arrangement of tags. For example,

```
<Unit name="Dacitic Vent Breccia">
<Age lowerBound = "-23.8" upperBound="-
   5.3">Miocene</age>
<Rock Name = "Dacite">
<Color>Light Medium Gray</Color>
<Texture>finely porphyritic</Texture>
<Constituent Role = "Inclusion">
   <Rock Name = "Granite">
   <Age lowerBound="-206" upperBound="-
      144">Jurassic</Age>
   </Rock>
   <Rock Name = "Granite">
   <Age lowerBound="-2500" upperBound="-
      543">Proterozoic</Age>
```

```
    </Rock>
    <Rock Name = "Rhyolite">
    <Age lowerBound="-206" upperBound="-
        144">Jurassic</Age>
    </Rock>
</Constituent>
<Genesis>Volcanic vent or throat</Genesis>
</Rock>
</Unit>
```

This is a very different document since there is no free text, only tags with attributes representing a described rock. One important aspect here is that `<Texture>` tags are embedded into the `<Rock>` tag. This implies that the textural description is tied to its `<Rock>` container; this is critical since the texture applies to the Dacite unit and not any other rock body.

This latter example is still readable by a human. The content of the XML document can be divided into small pieces of information easily handled by the computer. The class of softwares that transform XML documents into chunks of usable information is called **parser**, and many are available for free on the Internet. For anyone who writes softwares using XML, the parser will take care of all the details of loading, analysing and, more important, **validating** the XML document (we'll see what a valid document is in the next section). All this would need to be developed from scratch if any other format structure were chosen.

XML SCHEMAS

There are no predefined tags in XML. They, with their rules, must be defined by a group wishing to encode a specific domain. There are XML encodings to describe chemistry (CML), mathematic formula (MathML), geographic features (GML), poetry (XML Poetry) and even recipes (RecipeXML, formally known as DESSERT=Document Encoding and Structuring Specification for Electronic Recipes). Luckily, geoscience has not been left out; XMML (eXploration Mining Markup Language) is being developed in Australia to address mining industry requirements (see <http://www.ned.dem.csiro.au/XMML/>). Finally, our own effort is directed toward defining a mark-up language to describe geological maps.

As previously mentioned, a usable XML document must be both 'well formed' and 'valid'. A **well formed** document is a document that follows the basic rules that **all** XML-based markup documents must follow.

Rules include, but are not limited to:

1. When a tag is opened, it must be closed. For example, `<MYTAG>` must be followed by a `</MYTAG>`. (or if the tag does not carry any content, a valid shortcut is `<MYTAG/>`)
2. Tags must be nested, this means that tags must closed in the reversed order or opening. `<A><B></B></A>` is well formed while `<A><B></A></B>` is not.
3. Tags are case sensitive , so `<MyTag>` is not the same tag as `<MYTAG>`, therefore `<MYTAG>` . . . `</MyTag>` is not a well formed structure.

For an XML document to be **valid**, it must follow the domain specific rules. These rules can be defined using two mechanisms: 1) Document Type Definition (DTD) which is becoming used less often or restricted to some specialised tasks, and 2) XML Schema, which is itself an XML document that follows its own domain rules (the business of describing other domains). The goals of the schema creator are to distil from a domain a set of rules and encode them into an XML schema. The process is not straightforward and many schemas can produce similar XML documents describing the same domain but using a different approach. XML schema specification can be found from the W3C website <http://www.w3.org/XML/Schema>, but books on XML schemas, such as Duckett and others (2001), are great helpers.

A simple domain rule such as 'A Rock is made of at least one 'Mineral' can be translated in XML as 'The `<Rock>` tag must enclose one or more `<Mineral>` tags'. In XML Schema jargon, this says that `<Rock>` is a *complex* tag, enclosing a *sequence* of `<Mineral>` tags that appear at least once, up to an unspecified number of times. This rule indicates that the following XML document is **valid**:

```
<Rock>
<Mineral/>
<Mineral/>
</Rock>
```

while this one is not:

```
<Rock>
</Rock>
```

because at least one `<Mineral>` tag is required.

NADM CONCEPTUAL MODEL (NADM-C1)

For the *DITT*, we were lucky to have the NADM-specified domain for geologic maps already described by the DMDT. This team has worked for the last 2 years to develop the rules that describe geology and geological maps. The output of their effort is an extensive diagram, called NADM-C1 (North American Data Model Steering Committee's Data Model Design Team, 2003) showing principal geological features and how they relate to each other (the schematic for NADM-C1 (NADM Conceptual Model 1.0) is available from <http://geology.usgs.gov/dm/steering/teams/design>). Even if some aspects of the model are still being debated, a version 1.0 should be available soon.

The DITT must convert this information (see fig. 1 for a simplified example of a portion of the NADM-C1 diagram) into a format that is a suitable XML schema. As we pointed out, XML documents have their own sets of requirements and constraints. Therefore, decisions have to be made to ensure consistency in the conversion process. XML schema provide enough rope to hang any designer and no schematisation approach is 'better' that the other.

Many XML encoding styles exist. For example, some designers prefer the use of attributes,

```
<Rock Name = "Granite">
<Mineral/>
</Rock>
```

while some argue that we should avoid them and replace them by tags.

```
<Rock>
<Name>Granite</Name>
<Mineral/>
</Rock>
```

Both alternatives work, but they impose different constraints. Options must be evaluated and decisions must be made on the style to be used. This is critical to ensure consistency of the XML documents.

The simplified UML diagram of a portion of NADM-C1 (fig. 1) can be read as follows:

*A **CompoundMaterial** is a kind of **EarthMaterial,** which is a kind of GeologicConcept, that is composed of at least one other **EarthMaterial** but not limited to one. A GeologicConcept must be associated with least one Name, but can have more than one, and can also be associated with many **Descriptions,** but this is optional. A **CompoundMaterial** must be associated with one and only one **ConsolidationDegree**. A **Rock** is a kind of **CompoundMaterial;** therefore, it is also composed of other **EarthMaterials** and is also associated with a **ConsolidationDegree**. And since a CompoundMaterial is a GeologicConcept, a Rock is a GeologicConcept as well and must have at least one name, and potentially some descriptions. A Mineral is a kind of **EarthMaterial** (and is not a composition of **EarthMaterials**) . . .*

and so on.

We can tell from the diagram that a **Rock** is made of other **Rocks**, **Minerals**, **Fluids** and **Glasses**. All these components forming a **Rock** are optional, but at least one component must appear. For example, a quartzite is essentially made of only quartz mineral; a porous conglomerate is a mixture of rocks, mineral, possibly glass, voids and even fluids.

An XML schema translation of those rules is shown in figure 2. This figure has been made with XML Spy (<http://www.altova.com>); it simplifies the reading of

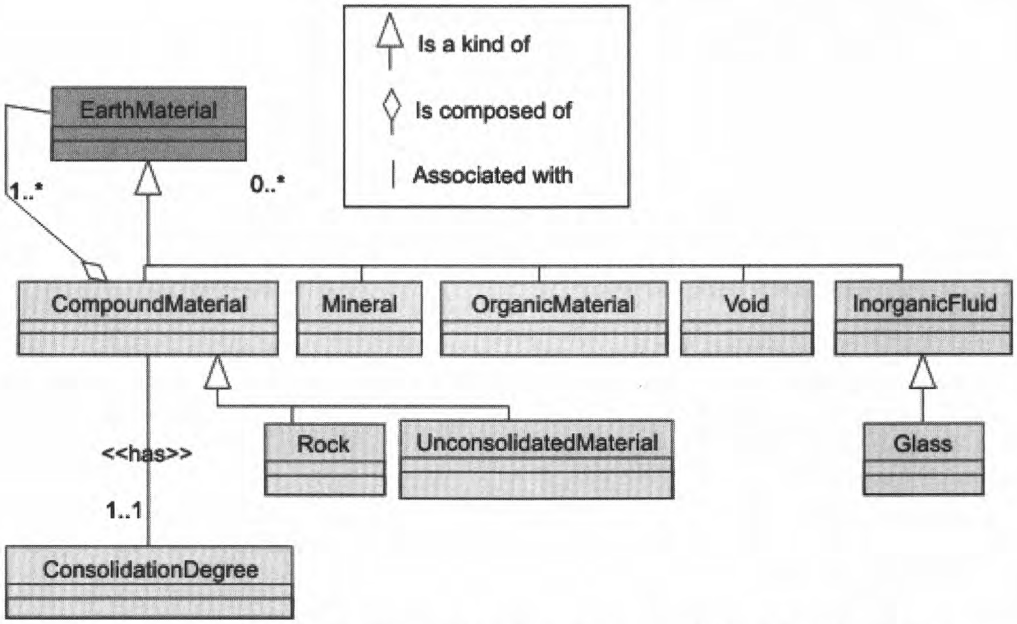
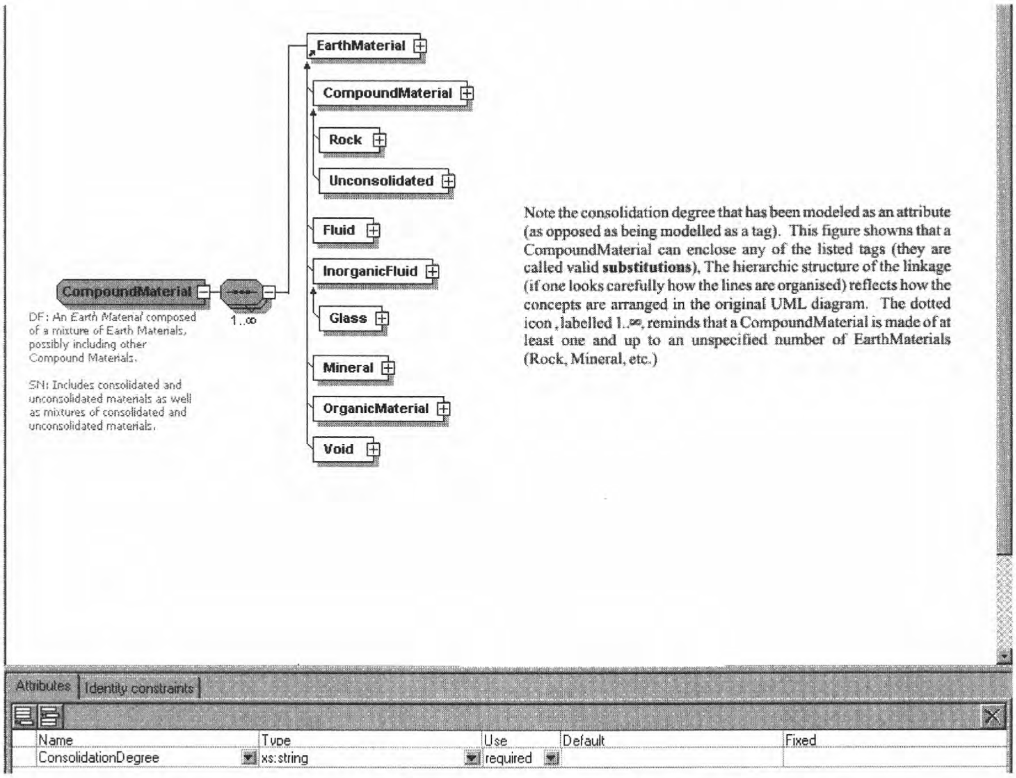


Figure 1. Excerpt of NADM-C1 Diagram showing a Compound Material made of other EarthMaterial.



DF: An *Earth Material* composed of a mixture of Earth Materials, possibly including other Compound Materials.

SN: Includes consolidated and unconsolidated materials as well as mixtures of consolidated and unconsolidated materials.

Note the consolidation degree that has been modeled as an attribute (as opposed as being modelled as a tag). This figure showns that a CompoundMaterial can enclose any of the listed tags (they are called valid **substitutions**), The hierarchic structure of the linkage (if one looks carefully how the lines are organised) reflects how the concepts are arranged in the original UML diagram. The dotted icon ,labelled 1..∞, reminds that a CompoundMaterial is made of at least one and up to an unspecified number of EarthMaterials (Rock, Mineral, etc.)

Name	Type	Use	Default	Fixed
ConsolidationDegree	xs:string	required		

Figure 2. Translation of figure 1 in an XML schema using XMLSpy.

XML schemas by representing tags and relations between them using an easy to understand schematic. For instance, the following document describing a compound material (remember, Rock is a CompoundMaterial) would validate against the XML schema depicted in figure 2:

```
<Rock ConsolidationDegree = “lithified”>
    <Name> Granite </Name>
    <Mineral>
        <Name> Biotite <Name/>
    </Mineral>
    <Mineral>
        <Name> Quartz <Name/>
    </Mineral>
</Rock>
```

But this next example would not be a valid NADM-C1 document because according to the NADM-C1 diagram, **Fluids** cannot contain **Minerals**.

```
<InorganicFluid>
    <Name>Water</Name>
    <Mineral>
        <Name>Salt</Name>
        <Name>Halite</Name>
    </Mineral>
</InorganicFluid>
```

The correct way to define it would be a mixture of Water and Mineral (not water *containing* mineral).

```
<CompoundMaterial ConsolidationDegree =
    “fluid”>
    <Name>Salty Water</Name>
    <InorganicFluid>
        <Name>Water</Name>
    </InorganicFluid>
    <Mineral>
        <Name>Salt</Name>
        <Name>Halite</Name>
    </Mineral>
</CompoundMaterial>
```

XML AS AN INTERCHANGE MECHANISM

The real benefit of XML encoding is the ease with which one can manipulate the document. In addition to the existing programming tools used to develop applications, another set of tools are available to transform XML documents from one schema to another. XSLT (eXtensible Stylesheet Language Transformations) is a specification to encode transformation rules (also called a ‘Style Sheet’) to convert an XML schema into another XML schema. XSLT is not restricted to XML transformation; in fact, an interesting application is the transformation of XML documents to a any text based document (for example, a series of SQL command). But we still have to keep in mind that XSLT has been designed for XML.

Transforming dynamically from one schema to another is the root of a technique called 'mediator-wrapper', where small softwares perform small translation tasks. A **wrapper** is a small application that translates a source dataset into an XML document or the other way around. The mediator takes several XML documents and manipulates them to create a new XML document suitable for another wrapper to handle and turn into something useful. Some mediators are merely transport mechanisms that do not transform the source XML but just convey it to a destination wrapper. Figure 3 shows such a mediator-wrapper architecture that transfers geological information from one database structure to another. The first wrapper transforms a subset of a database into a NADM-XML document, which is sent to another wrapper that extracts the information it needs to fill its own database. The first wrapper has no knowledge of the destination format, nor does the destination wrapper know anything about the source database. Each participating database must be able to translate between its own structure into NADM-XML and vice and versa.

Problems arise when we want to use the XML document to export a subset of a database. NADM-C1 is highly recursive; 'things' are described against other things, which in turn are also described using other 'things'. XML on the other hand is *sequential.* It is constructed from top to bottom, with tags nested into tags. Where does the document stop? Should it contain ALL related data? It could potentially export the complete database by following the linkage between geological 'things'. This forces us to think ahead on how to reference something outside a XML document. There are mechanisms in XML to point to something outside of the current document (Xpath, XPointer), but this mechanism assumes that the pointed elements are also inside the XML document. But in our case, the information might (and probably will) reside in a database. The actual schema, which is still in draft, is an encoding of the conceptual model, as if a document should contain all related data; the next step is to design those pointer mechanisms.

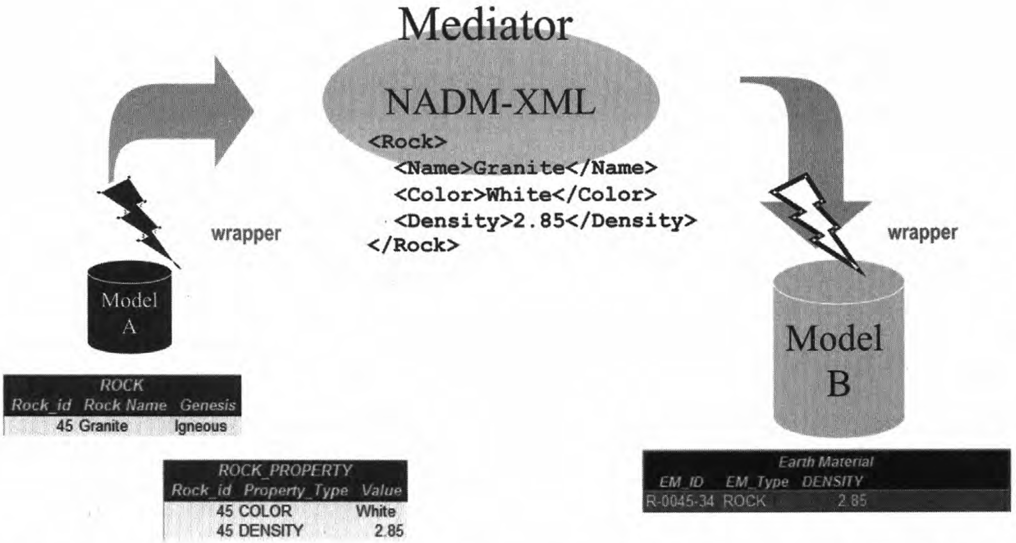
CONCLUSION

Encoding XML schema from the conceptual geologic data model is quite a challenge. Beside the substantial learning curve of schema design and the number of decisions that must be made to achieve a clean design, NADM-C1 is a complex model that reflects the complexity of geology. This complexity has an impact on the design of the schema. Lots of 'design patterns' and rules of thumb are available from the XML community. But as Alan Kay once said, "Simple things should be simple. Complex things should be possible" (Lipkie and others, 1982), and the payoff of this work should be a greater usability of geoscience and improved interoperability.

REFERENCES

Drewes, Harald, 1998, Geologic map of the Bartlett Mountain quadrangle, Pima and Santa Cruz Counties, Arizona: U.S. Geological Survey Miscellaneous Geologic Investigations Map I-2624, scale 1:24000.

Duckett, Jon, Griffin, Oliver, Mohr, Stephen, Francis, Wtokes-Rees, Ian, Williams, Kevin, Cagle, Kurt, Ozu, Nikola and Tennison, Jeni, 2001, Professional XML Schemas:



```
<Rock>
  <Name>Granite</Name>
  <Color>White</Color>
  <Density>2.85</Density>
</Rock>
```

ROCK

Rock_id	Rock Name	Genesis
45	Granite	Igneous

ROCK_PROPERTY

Rock_id	Property_Type	Value
45	COLOR	White
45	DENSITY	2.85

Earth Material

EM_ID	EM_Type	DENSITY
R-0045-34	ROCK	2.85

Figure 3. Scenario of an exchange of data between two databases of different structure using mediator-wrapper technology.

Birmingham, UK., Wrox Press ltd, 693 pages.

North American Data Model Steering Committee's Data Model Design Team, 2003, Proposed North American Standard Conceptual Data Model for Geologic Map Information. <http://geology.usgs.gov/dm/steering/teams/design/>

Lipkie, D. E., Evans, S. R., Newlin, J. K., and Weissman, R. L., Star graphics: An object-oriented implementation, *in* Proceedings, SIGGRAPH'82: Computer Graphics 16 (3), July 1982, pp. 115-124.

APPENDIX

Open source Movement:	<http://www.opensource.org/>
Leading XML portal:	<http://www.xml.org/>
W3C xml specifications	
XML:	<http://www.w3.org/XML/>
XSLT:	<http://www.w3.org/TR/xslt>
Schema:	<http://www.w3.org/XML/Schema>
XMML	<http://www.ned.dem.csiro.au/XMML/>
NADM-C1	<http://geology.usgs.gov/dm/steering/teams/design/NADM-C1.0/NADMC1_0.pdf>
XMLSpy:	<http://www.altova.com>

GEON: Toward a Cyberinfrastructure for the Geosciences—A Prototype for Geologic Map Integration via Domain Ontologies

By Bertram Ludäscher[1], Kai Lin[1], Boyan Brodaric[2], and Chaitan Baru[1]

[1]San Diego Supercomputer Center
University of California, San Diego
9500 Gilman Dr., MC 0505
La Jolla, CA 92093-0505
USA
Telephone: (858) 822-0864
e-mail: {ludaesch,klin,baru}@sdsc.edu

[2]Geological Survey of Canada
601 Booth Street
Ottawa, Ontario, K1A 0E8
Canada
Telephone (613) 992-3562
e-mail: brodaric@gsc.NRCan.gc.ca

ABSTRACT

When trying to combine different geologic maps, a number of interoperability challenges need to be overcome. We first provide an overview of those challenges and then briefly describe a mediator architecture devised to overcome them. We then focus on the problem of providing integrated access to a set of geologic maps from different state geological surveys, by defining a global view on the different local source schemas. Next we address the problem of content heterogeneity by defining an "integration ontology" to which the various local data content are mapped. The integration ontology consists of various sub-ontologies, such as one for geologic age (Poling, 1997), and several others relating to rock classification (chemical composition, texture, fabric, and genesis). The latter are derived from a recent proposal for rock classification (Struik and others, 2002). Based on the integration ontology, the prototype allows the user novel "concept-based" access and querying capabilities across the different geologic maps. This system is being embedded in the service-oriented data grid infrastructure under development in GEON.

INTRODUCTION

The Geoscience Network (GEON) is a collaborative NSF/ITR (National Sciences Foundation / Information Technology Research) project to create "cyberinfrastructure" for the Geosciences (GEON, 2003). GEON addresses the need in the geosciences to interlink and share multidisciplinary data sets in ways that allow researchers improved data access, data integration, and ultimately the construction of "scientific workflows" to combine data integration and analytical steps in a more seamless manner than is currently possible. The ultimate goal of GEON is to bring together heterogeneous scientific data and information to facilitate new ways of information integration and knowledge discovery.

The information technology (IT) research and development areas of GEON include data modeling and integration, grid systems development, and visualization. In the geoscientific component of GEON a set of scientific questions are centered around two test beds, the Rocky Mountain Region (uplift of Colorado Plateau), and the Mid-Atlantic Region (crustal (terrane) evolution). Data integration and mediation IT efforts are driven by and applied to the domain scientists' needs as exemplified by those test beds.

In this paper we describe one of the initial data integration efforts of GEON, the ontology-enabled map integration (OMI) system. Specifically, we describe the current prototype, which allows the user to query geologic maps provided by different state geologic surveys. The remainder of the paper is structured as follows: In the next section, we briefly recall the various levels of heterogeneity that present a challenge to data integration in general, followed by an overview of an extended mediator architecture that is used to overcome the interoperability challenges. Then, we elaborate on the ontology-enabled map integration prototype. Finally, we conclude and outline some future plans.

DATA INTEGRATION: OVERVIEW AND MEDIATOR ARCHITECTURE

Levels of Interoperability and Standards

When trying to share distributed, heterogeneous data, a number of technical challenges must be overcome; these can be roughly classified as described below (see also Sheth (1998) for a historical perspective). Consider, for example, two systems having data sets that should be made interoperable (Figure 1). One can employ standards and technologies to overcome the various kinds of heterogeneities, and to facilitate interoperability at different levels. At the systems level in Figure 1, we may find different operating systems (Linux, MS Windows, MacOS, ...), different data transport protocols (FTP or HTTP, which are built on top of a stack of internet protocols called TCP/IP), or higher-level protocols for discovery and interoperation of web services. Differences in system platforms and operating systems are usually overcome by standardizing protocols for data transport and remote service execution. For the latter, for example, one can employ web service descriptions (WSDL, 2001), which specify the input and output parameters of a web service. System level interoperability for GEON can also be achieved at the grid service level. Grid services extend the basic web service infrastructure and include additional features such as user authentication for secure data access. Apart from the generic issues of data access, transport, and remote execution, there are also a number of application specific system level issues, e.g., the choice and architecture of the mapping technology for rendering spatial information (server-side, client-side, mixed).

At the syntactic level, we consider heterogeneities such as different data file formats, e.g. SHP (ESRI shape files) and DXF (AutoCAD drawings). XML, the Extensible Markup Language (XML, 2000) provides a simple and very flexible syntax for structuring many kinds of data and metadata to enable their exchange. Defining such a new structure in XML syntax can be done in different ways. For example, one can provide an XML Document Type Definition (DTD) or an XML Schema definition (XML Schema, 2001) to specify the allowed nesting structure and (in XML Schema) the data types of XML elements. This not only yields a data exchange syntax but also prescribes a schema for the exchanged data. Additional semantics such as domain specific integrity constraints have to be encoded by other means. The Resource Description Framework (RDF, 2003) can be seen as an XML dialect for encoding labeled, directed graphs, in particular ontologies (see below). For querying databases, query languages such as SQL (for relational databases) or XQuery (for XML databases) are used, each of which come with their own syntax for query expressions. Differences at the syntactic level are usually resolved either by adhering to a standard or by using format converters that can translate from one format to another.

At the schema level, heterogeneities can exist because

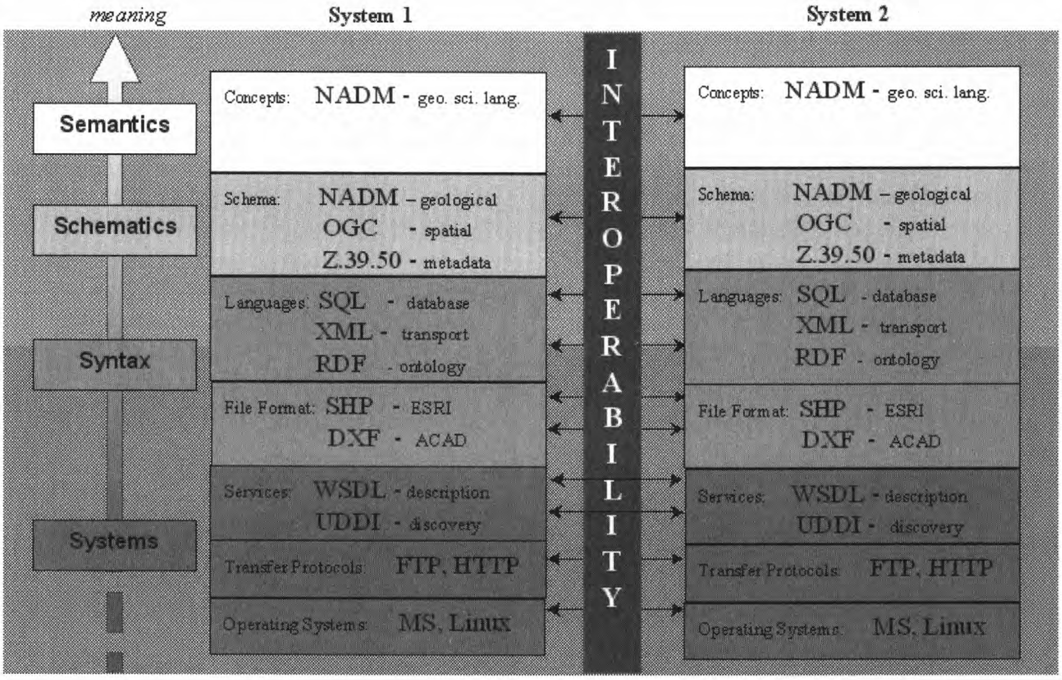


Figure 1. Levels of interoperability and standards.

the same (or at least similar) data can be represented using vastly different schema structures (even when the same file format or syntax is used). For example, two datasets may be organized in different ways across two relational databases, i.e., the table and column structure may be very different although the content (at the conceptual level) of the databases may be very similar. Similarly, for XML databases, different DTDs or XML Schemas can be used to describe the same data. To overcome schema level heterogeneities, we can again apply two approaches, schema standardization or schema transformation. An example of the former is Z39.50, which provides a syntax and schema for querying digital library collections. The Open GIS Consortium (OGC, 2003) and the Federal Geographic Data Committee (FGDC, 2003) provide a number of standards that cover syntactic and schema aspects, as well as controlled vocabularies. The North American Data Model (NADM, 2003) includes a comprehensive data model and schema suited for geologic maps and related information. For the latter, i.e., schema transformation, database query languages in general and XQuery in particular provide powerful means to express complex queries and transformations. Thus (XML) query languages play an important role in database mediators (see below).

Finally, at the semantic level, we consider issues such as differences in terminology, different classification schemes (such as for rock types), and differences in the definition of and constraints for the various concepts that are relevant to the data sets being integrated. The main approach for reconciling semantic heterogeneities is the use of agreed-upon ontologies, which in their simplest form provide a controlled vocabulary with more or less formal descriptions of the pertinent concepts. In more sophisticated forms ontologies include formalizations (often through logic formulas) of properties of concepts and "inter-dependencies" of concepts. A prominent emerging standard for ontologies is the Ontology Web Language (OWL, 2003), which comes in three increasingly expressive variants: OWL Lite, OWL DL, and OWL Full. OWL is also an interesting example of how several interoperability levels and standards may be intertwined: for example, OWL DL builds upon the RDF model and syntax which in turn is usually denoted in XML syntax.

Mediator Architecture

Database mediator systems can be used to provide uniform access to distributed heterogeneous data sets, and thereby overcome a number of the interoperability challenges mentioned above. Figure 2 depicts a typical mediator architecture, in which a number of local data sources are "wrapped" as XML sources and subsequently combined into an integrated global view G. Thus a client application or end user is provided with the illusion of

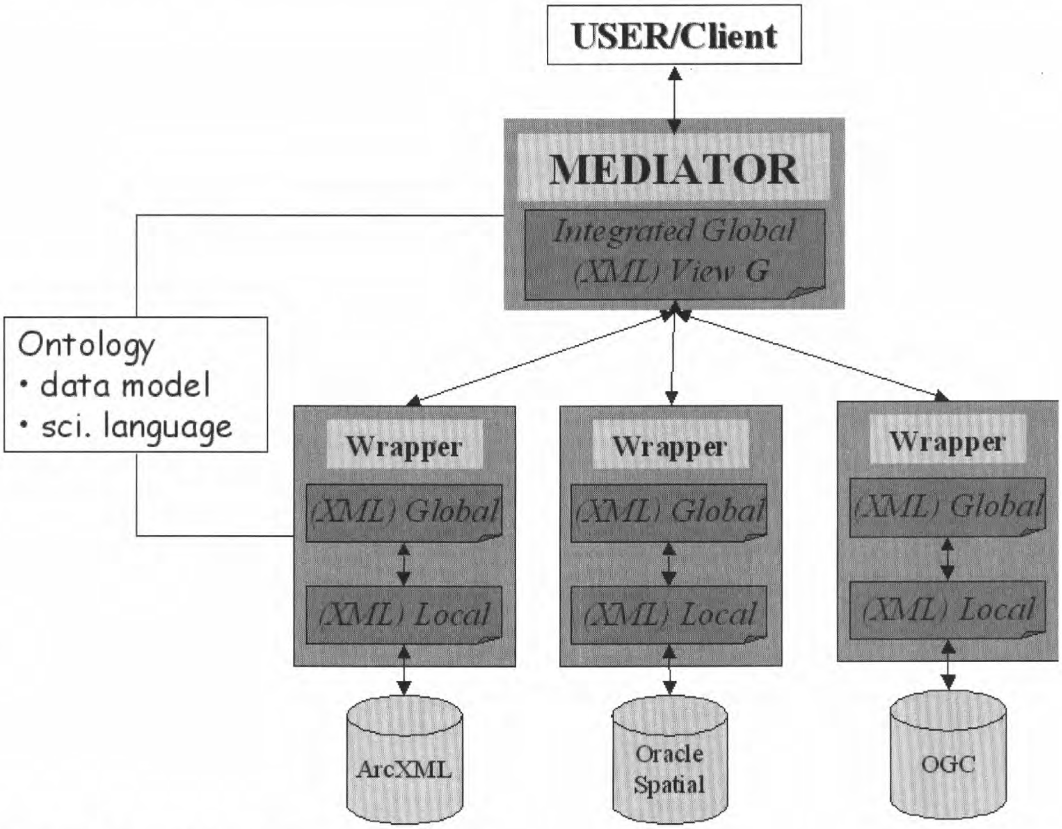


Figure 2. Extended mediator architecture.

querying a single, integrated (or global) database G.

The source wrappers not only provide a uniform syntax, but also reconcile system aspects, e.g., by means of a unified data access and query protocol. In a conventional relational or XML-based mediator system, interoperability is facilitated at the structural level. That is, differences in schema can be reconciled by corresponding schema transformation as part of the view definitions for the global view G. However, terminological differences or other semantic differences are not adequately handled at the purely structural/XML level. To this end, source schema and contents can be registered to an ontology, which encodes additional "knowledge" about the registered concepts. As we shall see in the next section, by "ontology-enabling" the system in this way, one can evaluate high-level queries over concepts that are not directly in the source databases, yet indirectly linked via the ontology.

THE ONTOLOGY-ENABLED MAP INTEGRATION SYSTEM

Figure 3 shows an end-user's view of the current ontology-enabled map integration prototype (OMI, 2003). Nine different spatial data sources, i.e., geological maps from various state geological surveys, are made interoperable in such a way that the user can seamlessly query across the different data sets, and even view the data through "conceptual-level glasses". By this we mean that once a source has been "semantically registered" relative to an existing ontology (such as the rock classification system of Struik et al., 2002), we can use that ontology's conceptual entities and relations to query the data set.

On the left in Figure 3 we see a collection of geologic maps from a number of state geological surveys. The query forms on the right of the depicted windows allow the user to query for regions having a specified geologic age and/or rock type. On the upper right, the results of a conventional query with GEOLOGIC-AGE='Paleozoic' are shown. Note that the system finds only very few regions, since the information that the Paleozoic Era contains periods such as Permian and Carboniferous, etc., is not "known" to the system. In contrast, on the lower right, we have "turned on" a geologic age ontology (Poling, 1997), and a much larger set of data is now found. Here, we have used a technique called "concept expansion" that replaces a query term such as 'Paleozoic' by all suitable "sub-concepts" (here the Periods, etc., belonging to 'Paleozoic') in order to retrieve all relevant data.

The Use and Role of Ontologies

In information integration systems (Figure 2), ontologies can be used to provide information at the level of conceptual models and terminologies, thereby facilitating conceptual-level queries against sources, and resolving some of the semantic-level heterogeneities between them. In our system (Figure 3), the rock classification ontology and geologic age ontology are used as a global view for registering data sets and processing queries. When a data set is registered to an ontology, a mapping from the data set to the selected ontology is generated. The registration process consists of the following three steps:

1. Select classes in the system ontology repository to register this data set: e.g., select the time scale and/or rock classification systems to be used as the global ontologies into which the data structure and contents are to be mapped;
2. Select columns for each selected class for populating virtual objects in these classes: e.g., map the columns in the data set to specific ontologies, thus indicating that a column's contents (their range) are mapped to an ontology, such as mapping a lithology column onto a rock classification ontology;
3. Select the populating methods or populate manually: map the column contents to classes in the ontology manually or semi-automaitcally using word-matching or other provided techniques, e.g., map "granite" from a lithology column to "Granite" in the ontology.

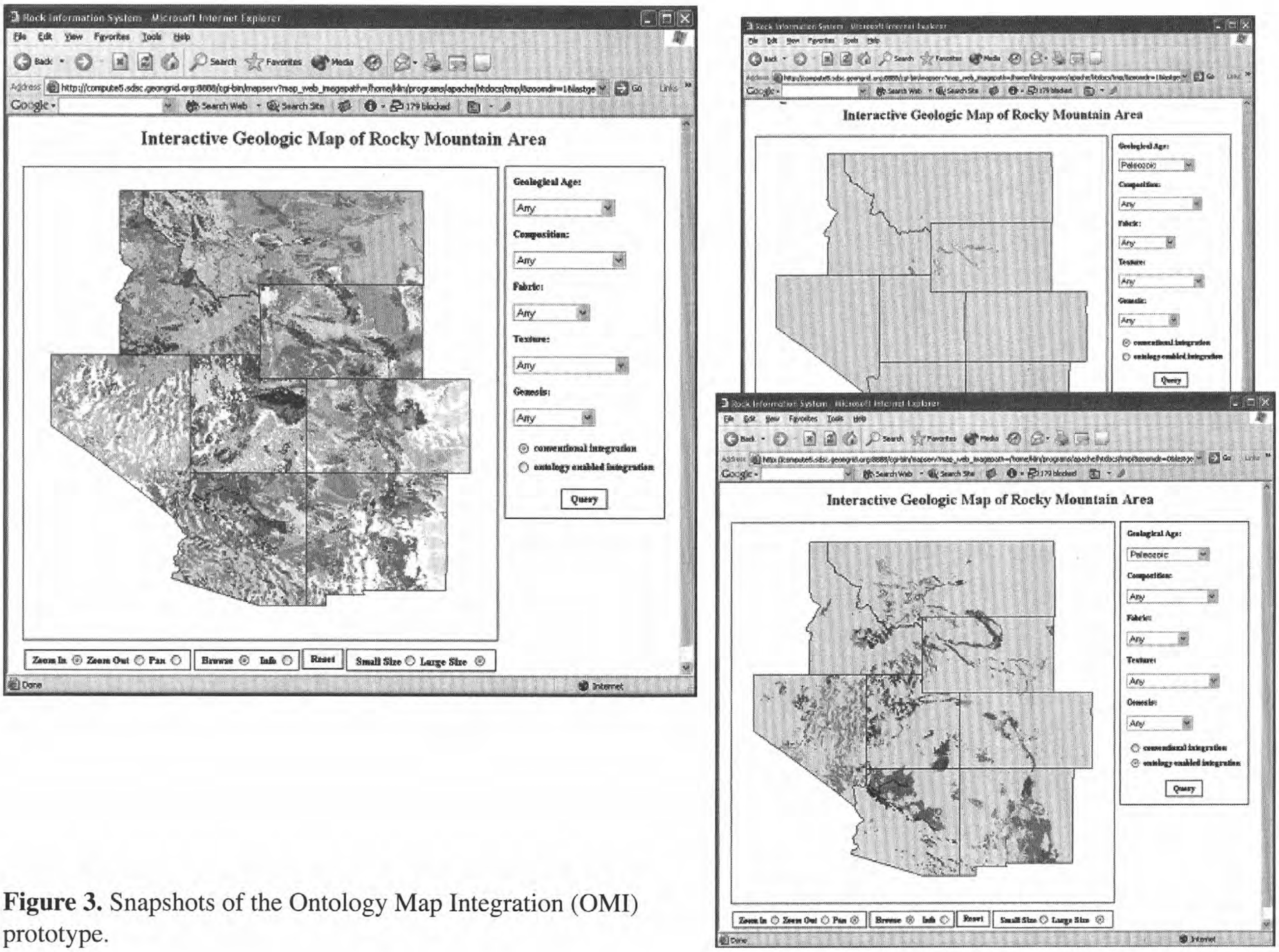
However, before such mapping can occur the sources' local data schemas must first be registered. For example, in our implementation we used the following two schemas for the Arizona and Idaho data sets:

Arizona—(AREA, PERIMETER, AZ_1000_, AZ_1000_ID, GEO, PERIOD, ABBREV, DESCR, D_SYMBOL, P_SYMBOL)

Idaho—(AREA, PERIMETER, ID_500_, ID_500_ID, FORMATION, UNIT_NAME, ROCK_TYPE, ERA, SYSTEM, SERIES, LITH1, LITH2, LITH3, LITH4, LITH5, LITH6, LITH7, LITH8, LOCATION1, LOCATION2, COMMENTS, IDCARB, IDK, IDBASE, IDFAM, IDPHOS, IDSG, IDBATHAB, LITHA, LITH_FORM, PERIOD, D_SYMBOL, P_SYMBOL, LITH_MAJOR, LITH_MINOR, LITHOLOGY, AGE, IDLITH)

After these steps, wrappers are created for the registered data sets. Each wrapper uses the mappings between the data source and ontology to translate queries from the global ontology to the local schema, and also to translate content from the local schema to the global ontology. As explained above, the system can automatically use the subclass relation to expand concept queries when required.

Note that although all system-registered ontologies can be considered as conceptual-level query mechanisms, the system can suggest suitable ontologies based on, first, the user's choice of data sets and, second, the sources' schema information.



Interactive Geologic Map of Rocky Mountain Area

Geological Age: Any
Composition: Any
Fabric: Any
Texture: Any
Genesis: Any
● conventional integration
○ ontology enabled integration
Query

Zoom In ● Zoom Out ○ Pan ○ | Browse ● Info ○ | Reset | Small Size ○ Large Size ●

Interactive Geologic Map of Rocky Mountain Area

Geological Age: Paleozoic
Composition: Any
Fabric: Any
Texture: Any
Genesis: Any
● conventional integration
○ ontology enabled integration
Query

Interactive Geologic Map of Rocky Mountain Area

Geological Age: Paleozoic
Composition: Any
Fabric: Any
Texture: Any
Genesis: Any
○ conventional integration
● ontology enabled integration
Query

Zoom In ○ Zoom Out ○ Pan ● | Browse ● Info ○ | Reset | Small Size ○ Large Size ●

Figure 3. Snapshots of the Ontology Map Integration (OMI) prototype.

The Architecture of the Prototype

Figure 4 shows the system architecture: the system consists of a http server, a query mediator, and a map server (MapServer, 2003). When a request is received by the web server, the system generates a query against the global ontology, and then sends the query to the mediator; the mediator decomposes the query into several subqueries, and sends each of them to its target database. Then a mapfile is created based on these query results, and finally the map server renders a map according to the mapfile, and sends the map back to the user. Note that a map can contain several remote layers from other web map servers.

If a query, for example, asking to show rocks of Cenozoic age is received, the system takes the following steps to process the query:

1. Concept expansion: gets all the subclasses of the queried class (not shown in Figure 4).
2. Query rewriting: generates new queries to find formations against the two virtual tables by using the subclasses in the set found in step 1;
3. Map rendering: renders a map based on the query results of step 2 and predefined colors.

CONCLUSIONS AND FUTURE PLANS

We have described our current ontology-enabled prototype for integrating geologic maps from different sources. Syntactic and structural differences are overcome by traditional schema integration and database mediation techniques. In addition, "semantic mediation" and conceptual-level queries are supported by registering source data sets to domain ontologies such as for geologic age, rock type classification, etc.

On the systems side, we are adding commercial mapping technology (ESRI) in addition to the current open source technology (MapServer, 2003). Moreover, we will also "grid-enable" the application, i.e., use Grid standards for data access and querying. At the level of ontologies, we are working on a "3rd-party registration mechanism", that will allow the user to register a data set relative to one ontology (rock type ontology A), and then query the data set using another ontology (rock type ontology B); cf. (Lin and Ludäscher, 2003; Bowers and Ludäscher, 2003). This is only possible by having a "mediation engineer" devise a so-called articulation ontology that maps concepts between ontologies (such as A and B). We have already conducted preliminary studies in this

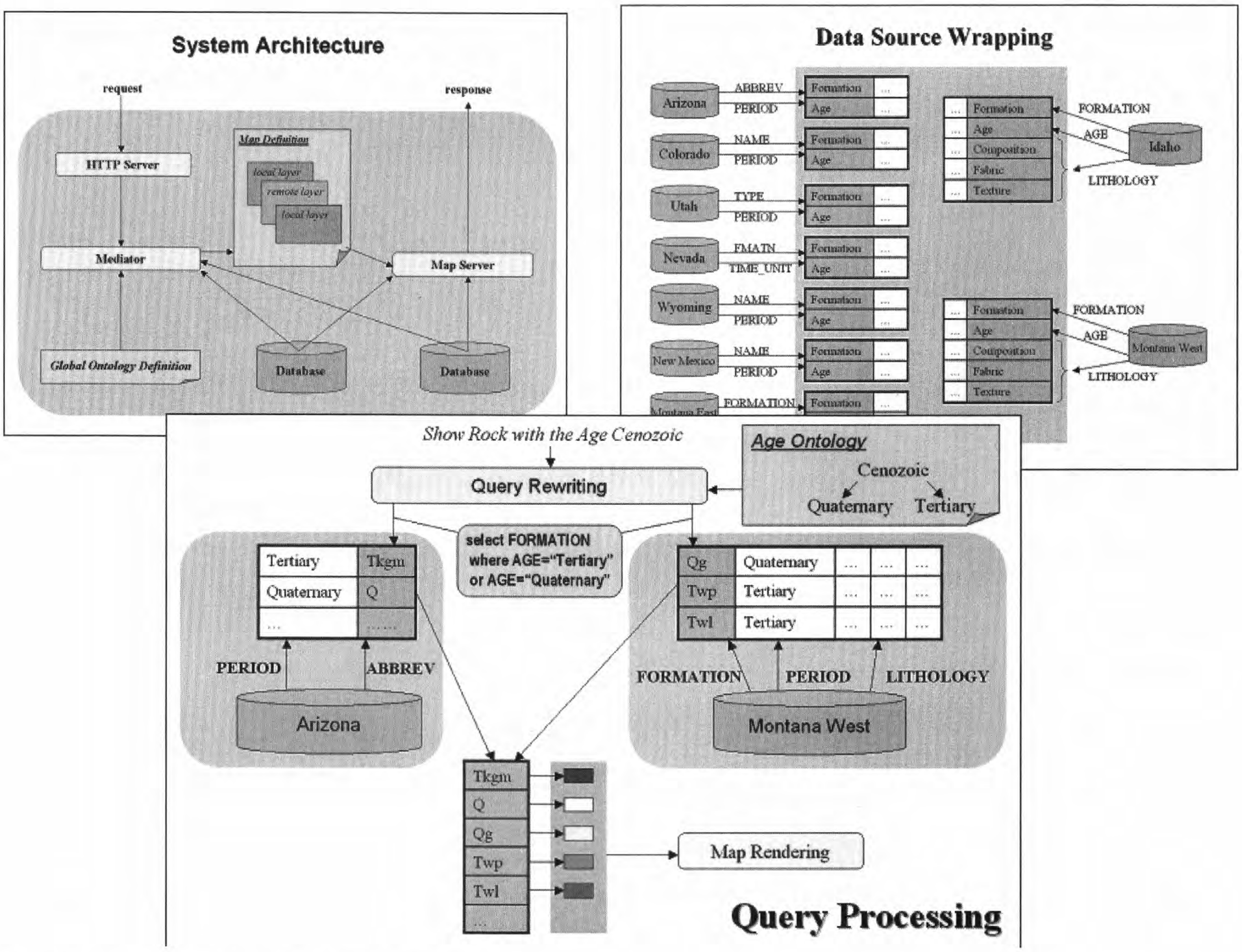


Figure 4. System architecture of the prototype.

direction, which employ OWL to map concepts between different ontologies (for example, the rock type classifications of the British Geological Survey and Struik and others, 2002). Ultimately we are interested in embedding applications such as the one described here into a GEON workflow environment that will allow the user to combine data integration steps (e.g., geologic map integration) with analytical steps (e.g., rock classification) to form a high-level "scientific workflow".

REFERENCES

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P., 2003, The description logic handbook—Theory, implementation and applications, Cambridge University Press.

Bowers, S. and Ludäscher, B., 2003. Towards a generic framework for semantic registration of scientific data, in Semantic web technologies for searching and retrieving scientific data (SCISW), <http://www.sdsc.edu/~ludaesch/Paper/scisw03-seek.pdf>.

FGDC, 2003, Federal Geographic Data Committee, <http://www.fgdc.gov>.

GEON, 2003, The Geosciences Network, <http://www.geongrid.org>.

Lin, K., and Ludäscher, B., 2003, A system for semantic integration of geologic maps via ontologies, in Semantic web technologies for searching and retrieving scientific data (SCISW). <http://www.sdsc.edu/~ludaesch/Paper/scisw03-geon.pdf>.

MapServer, 2003, University of Minnesota, <http://mapserver.gis.umn.edu/>.

NADM, 2003, North American digital geologic map data model, <http://geology.usgs.gov/dm/steering/teams/design/>.

OGC, 2003, Open GIS Consortium, <http://www.opengis.org>.

OMI, 2003, Ontology-enabled map integration prototype, <http://kbis.sdsc.edu/GEON/map-integration.html>.

OWL, 2003, OWL Web ontology language reference, W3C candidate recommendation 18 August 2003, <http://www.w3.org/TR/owl-ref/>.

Poling, J., 1997, Geologic ages of earth history, <http://www.dinosauria.com/dml/history.htm>.

RDF, 2003, Resource Description Framework (RDF), <http://www.w3.org/RDF/>.

Sheth, A., 1998, Changing focus on interoperability in information systems: From system, syntax, structure to semantics, interoperating Geographic Information Systems, pp. 5–30, Kluwer.

Struik, L., Quat, M., Davenport, P., and Qkulitch, A., 2002, A preliminary scheme for multihierarchical rock classification for use with thematic computer-based query systems: Geological Survey of Canada, <http://www.nrcan.gc.ca/gsc/bookstore/free/cr_2002/D10.pdf>.

WSDL, 2001, Web Services Description Language (WSDL) 1.1, W3C Note 15 March 2001, <http://www.w3.org/TR/wsdl>.

XML, 2000, Extensible Markup Language (XML) 1.0 (Second Edition), W3C Recommendation 6 October 2000, < http://www.w3.org/TR/REC-xml>.

XML Schema, 2001, XML Schema Part 0: Primer, Part 1: Structures, Part 2: Datatypes, W3C Recommendation 2 May 2001, <http://www.w3.org/TR/xmlschema-0/>.

Delaware Inland Bays Shoreline Extraction Using Landsat 7 Satellite Imagery

By Lillian T. Wang

Delaware Geological Survey
University of Delaware
Delaware Geological Survey Building
Newark, DE 19716
Telephone: (302) 831-1096
Fax: (302) 831-3579
e-mail: lillian@udel.edu

INTRODUCTION

As part of a larger study to map ground-water discharge areas in surface water, Landsat 7 Enhanced Thematic Mapper satellite imagery was used to delineate the coastline for Rehoboth and Indian River Bays, Delaware (fig. 1). Definition of a shoreline was critical in order to isolate pixels that included only open water in the bays. Automated, unbiased methods to extract the shoreline from January 29, 2000 (fig. 2), and February 19, 2002, Landsat images were evaluated. Existing vector shorelines from, for example, the United States Geological Survey (USGS) Digital Line Graphs (DLGs) or the Delaware Land-Use/Land Cover (LULC) data were not used because both vector shorelines are based on data sets that precede the Landsat image acquisition date, both vector shorelines have different resolutions than the Landsat image, and georeferencing errors needed to be minimized during the analysis.

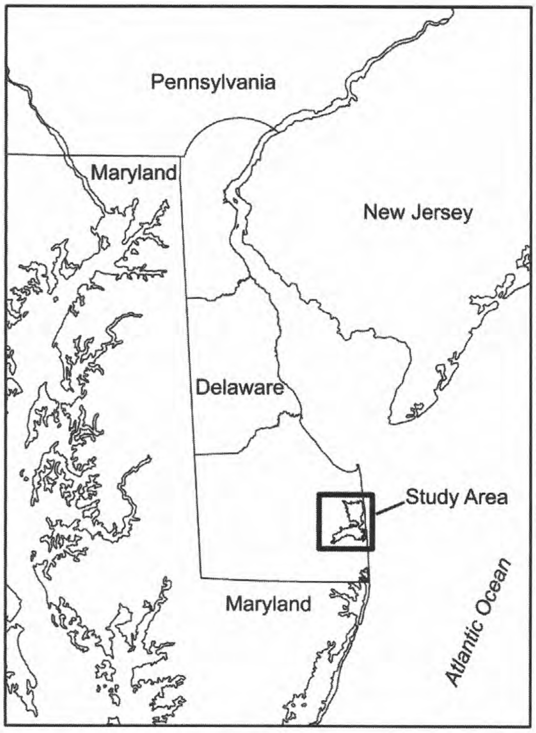


Figure 1. Location of Inland Bays, Delaware.

Several unsupervised classifications were performed in the Environment for Visualizing Images (ENVI) 3.6 image processing software, using individual bands and band combinations to determine the best spectral wavelength or combination of wavelengths for coastline delineation. The classification of the January Landsat image was initially troubling because known open water in the bays was not classified correctly. It was determined that these misclassified areas were caused by the presence of ice on the bays (fig. 3).

IMAGE PROCESSING STEPS

The near-infrared (NIR) and mid-infrared (MIR) wavelength bands normally provide high contrast between land and water (Jensen, 2000). An unsupervised classification of the NIR band classified most of the Inland Bays as water, but there were problem areas such as in the northwestern corner of Rehoboth Bay, shown in figure 4a. MIR wavelength bands were more successful in differentiating land and water because they are sensitive to moisture content (Gibbons and others, 1989; Schneider and Mauser, 1996; Jensen, 2000). However, obviously misclassified pixels still remained in the open water areas (fig. 4b).

Analysis of all bands seemed to be the next logical step. Visible, NIR and MIR bands were stacked and analyzed using tasseled-cap transformation, a useful tool for compressing spectral data into a few components associated with physical scene characteristics (Crist and Cicone, 1984). The tasseled cap consists of three primary factors: "Brightness" (soil brightness index), "Greenness" (green vegetation index) and "Wetness" (related to soil moisture

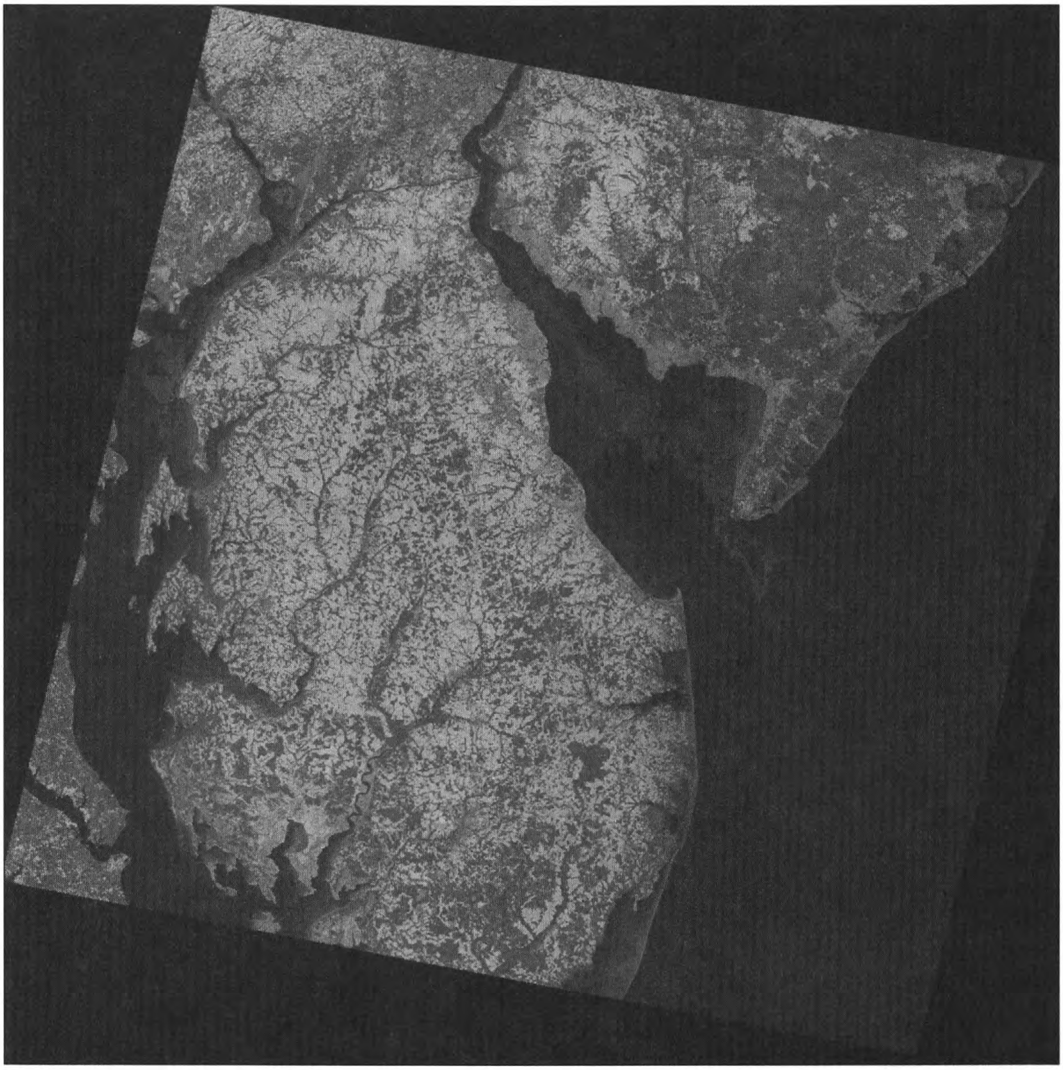


Figure 2. Landsat image, Path 014 Row 33, January 29, 2000.

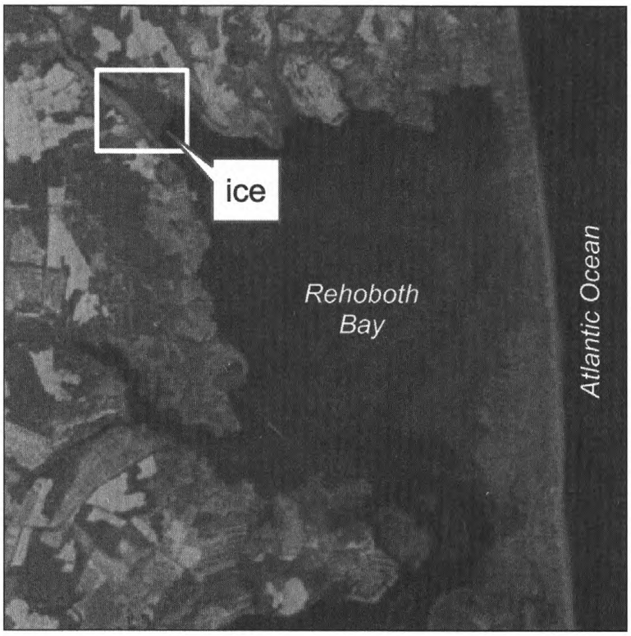


Figure 3. Ice located in the northwestern corner of Rehoboth Bay.

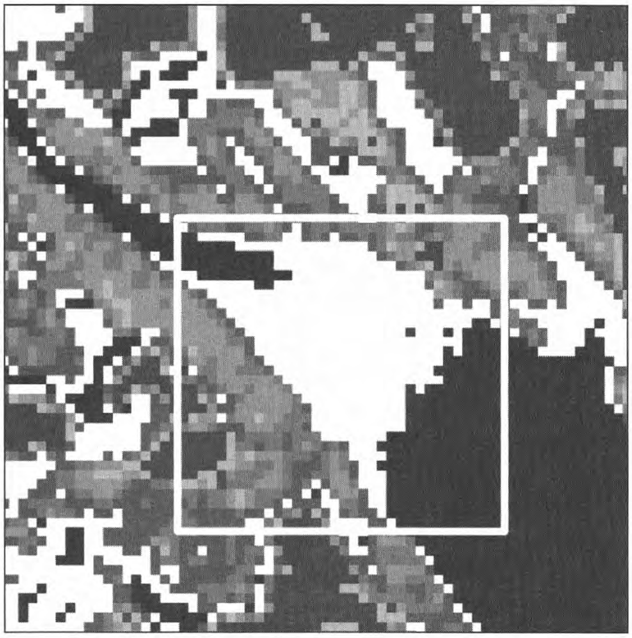


Figure 4a. Unsupervised classification of the NIR band (dark pixels = water).

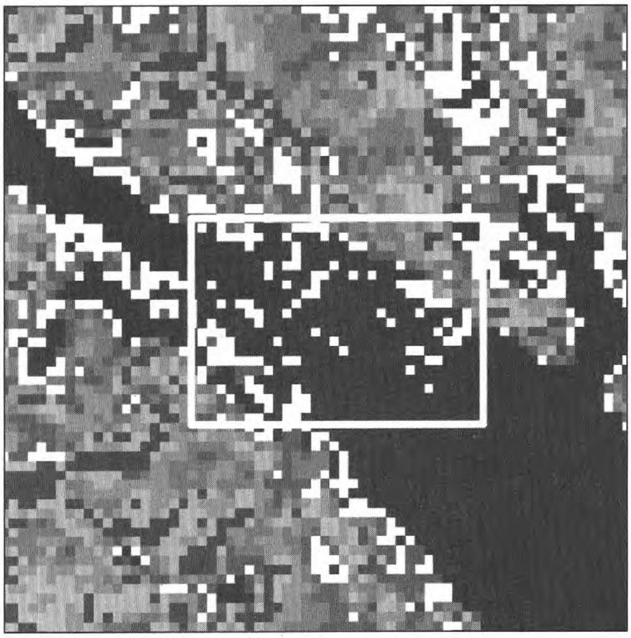


Figure 4b. Unsupervised classification of the MIR bands.

content). Although used mainly for vegetation studies, tasseled-cap transformation can separate urban, water, and wetland classes (Jensen, 1996).

This transformation resulted in some areas of open water still classified as land, and vice-versa, in the January 29, 2000 image. However, the same process run on the February 19, 2002, image, resulted in very good delineation between land and water. Meteorological data indicated air temperatures at or below freezing several days before the January image was taken. It has been determined that these misclassified areas were ice, and further processing was required to delineate the shoreline.

Classified images often exhibit a lack of spatial coherency. A post-classification technique called *sieve classes* is used to solve the problem of isolated pixels occurring in classified images, leaving unclassified black pixels. The *clump classes* technique is then used to clump adjacent similar classified areas, where unclassified pixels are reclassified (fig. 5). Results were converted to an ENVI polygon vector layer, and the water class was exported to an ESRI shapefile. Minor problem areas still existed, and the results were manually edited in ArcGIS using a false-color composite as a base map for visual aid.

SUMMARY

Shorelines from older DLG and LULC datasets were not used to delineate a coastline for Rehoboth and Indian River Bays because they were derived from data several years older and have a different spatial resolution than the satellite image. The final, post-processed shoreline is specific to the January 29, 2000, Landsat 7 image, and excludes land pixels as seen in the DLG (figure 6) and LULC shoreline datasets. The tasseled-cap transform was sufficient for shoreline extraction, but the mid-infrared band combination provided the best delineation of the coastline when ice was present on the surface water. Problems encountered in this study were found to be image-specific, due to cold weather conditions in January forming ice on the Inland Bays, Delaware.

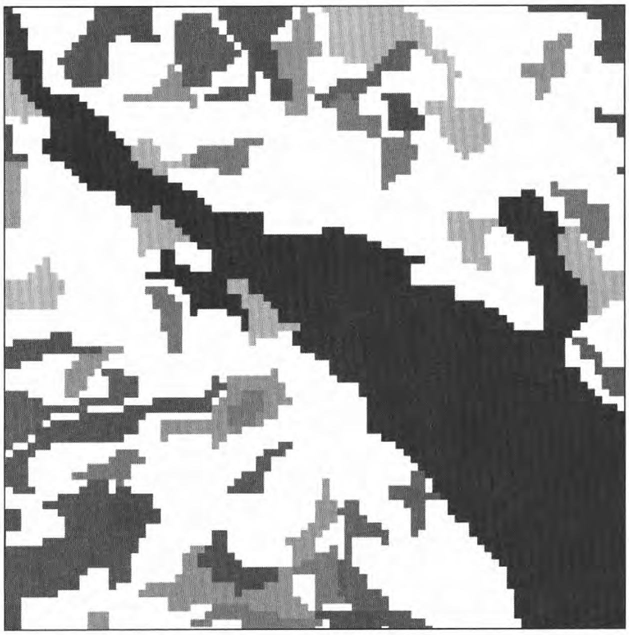


Figure 5. Result after *sieve* and *clump classes* run on figure 4b.

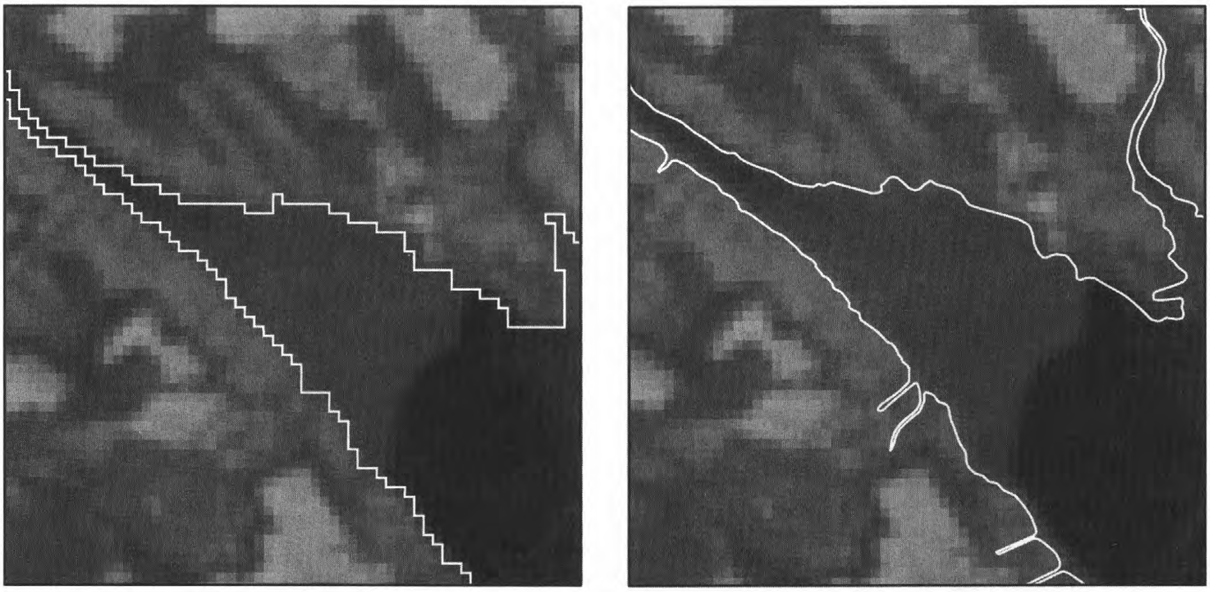


Figure 6. Final shoreline (left) compared to USGS DLG 1993 shoreline (right).

ACKNOWLEDGMENT

This research is funded by the US EPA – Science to Achieve Results (STAR) Program under grant R826945.

REFERENCES

Crist, E.P., and Cicone, R.C., 1984, A physically-based transformation of Thematic Mapper data - the TM Tasseled Cap: IEEE Transactions on Geosciences and Remote Sensing, GE-22, p. 256-263.

Gibbons, D.E., Wekelic, G.E., Leighton, J.P. and Doyle, M.J., 1989, Application of Landsat Thematic Mapper data for coastal thermal plume analysis at Diablo Canyon: Photogrammetric Engineering and Remote Sensing, v. 55, p. 903-909.

Jensen, J.R., 1996, Introductory digital image processing, (2nd ed.ed.): Upper Saddle River, NJ, Prentice Hall p. 184.

Jensen, J.R., 2000, Remote sensing of the environment - An earth resource perspective: Upper Saddle River, NJ, Prentice Hall, pp. 194, 382.

Schneider, K., and Mauser, W., 1996, Processing and accuracy of Landsat Thematic Mapper data for lake surface temperature measurement: International Journal of Remote Sensing, v. 17, p. 2027-2041.

The Environment for Visualizing Images (ENVI) – Research Systems, Inc., 4990 Pearl East Circle, Boulder, CO 80301 USA, (303) 786-9900, <http://www.rsinc.com>.

ESRI, ArcGIS – Environmental Systems Research Institute, Inc., 380 New York St., Redlands, CA 92373-8100 USA, (909) 793-2853, <http://www.esri.com>.

Applying Surficial Geologic Mapping to Geologic Hazards Mapping, Nez Perce County, Idaho

By Loudon R. Stanford,[1] Kurt L. Othberg,[1] Bill Reynolds,[2]
and Roy M. Breckenridge[1] Idaho Geological Survey

[1]Idaho Geological Survey
Third Floor, Morrill Hall
University of Idaho
Moscow, ID 83844-3014
Telephone: (208) 885-7479
Fax: (208) 885-5826
e-mail: stanford@uidaho.edu, roybreck@uidaho.edu, othberg@uidaho.edu

[2]Nez Perce County
1230 Main street
Lewiston, ID 83501-0896
Telephone: (208) 750-2055
e-mail: BillReynolds@co.nezperce.id.us

INTRODUCTION

The Idaho Geological Survey's long-term geologic mapping plan is designed to serve the natural-resource base and the growing population of Idaho. However, utilization of geologic mapping by counties and cities has lagged even as maps have become available. During the 1990s, this underutilization began to change as more counties and cities began to use geographic information system (GIS) databases for their planning and decision making. Since 1999, the Idaho Geological Survey has coordinated new surficial geologic mapping with development of a geologic hazards component of Nez Perce County's newly implemented ArcView GIS. The Idaho Geological Survey cooperated with the commissioners and planning department of Nez Perce County to identify critical areas of need. Because of recurrent landslides, the county had retained a geotechnical consultant and was beginning to understand the utility of basic geologic mapping to planning, zoning, and permitting.

SURFICIAL GEOLOGY

The surficial mapping used for Nez Perce County was from STATEMAP projects in the western and central parts of the county near the city of Lewiston and along the U.S. Highway 12 corridor. These areas are located in northern Idaho near the boundary between the Columbia Plateau and the Northern Rocky Mountains. The Clearwater River valley is steep-sided and large landslides are common. The surficial geologic map shows units and symbols that characterize geomorphic processes and their potential as geologic hazards.

GEOTECHNICAL TERRAIN UNITS

The county needs geologic mapping to help it identify geologic-hazards and locate areas that require site-specific geotechnical studies. The surficial geologic map is vital information; however, county decision makers are unable to directly translate the geologic units to practical engineering categories. The county's geotechnical contractor interpreted the engineering properties and material characteristics of the geologic units into "geotechnical terrain units" for the county. Each geotechnical terrain unit (GTU) includes a description of its capabilities for the following categories: slope, ground water, erosion, soils, earthwork, roadways, foundations, septic systems, and whether or not a site-specific study should be required.

CONCLUSION

The collaboration by the county, its geotechnical contractor, and the Idaho Geological Survey through its STATEMAP project, was ideal in terms of timing. Because of costly damages from landslides, Nez

Perce County was prepared to make landslide hazard interpretations a priority in its budget when the Idaho Geological Survey began STATEMAP surficial geologic mapping in the area. As the survey's mapping data became available following each STATEMAP contract period, the county's geotechnical contractor interpreted the GTUs on the basis of the surficial geology, and each data set was imported into an ArcView coverage for the county planning department. The Idaho Bureau of Disaster Services and the State Mapping Advisory Committee have been supportive of this project and similar applications in other counties.

Idaho counties vary in their capability and willingness to incorporate geologic mapping and geotechnical interpretations into their planning, zoning, and permitting procedures. The successful collaboration between the Idaho Survey and Nez Perce County is, however, serving as a model for increased utilization of geologic mapping elsewhere in Idaho.

REFERENCES

Howard, T. R., 2003 Geotechnical Hazard Mapping in Nez Perce County, Idaho: Unpublished mapping by the Idaho Geological Survey and Howard Consultants

Howard, T. R., 2003 Geotechnical Hazard Mapping in Clearwater County, Idaho: Unpublished mapping by the Idaho Geological Survey and Howard Consultants

Creating a Shaded-Relief Geologic Map Using World Construction Set and Adobe Illustrator Software

By Loudon R. Stanford

Idaho Geological Survey
Third Floor, Morrill Hall
University of Idaho
Moscow, ID 83844-3014
Telephone: (208) 885-7479
Fax: (208) 885-5826
e-mail: stanford@uidaho.edu

INTRODUCTION

One of the continuing goals of geologic mappers and map makers is to find better ways to show geology on a map. Shaded terrain, when combined with colored geology polygons, makes a geologic map more visually appealing and easier to use. This poster demonstrates one approach to creating such a map.

WORLD CONSTRUCTION SET

The shadedrelief for the *Surficial Geologic Map of the Wood River Valley Area, Blaine County, Idaho*, was created in the visualization software World Construction Set 5. World Construction Set (WCS) uses fractal geometry and material controls to aid in the creation of 3D landscapes. Geometric features that look similar, no matter the scale, have fractal properties. A coastline has fractal properties and so does the topography of the earth'ssurface. When combined with digital elevation data, the fractal math applied by WCS "fillsin" or adds detail as the scale of the scene gets larger, or as you zoom in. WCS also allows the user to manipulate the color and texture of the ground, water, vegetation, and sky. In addition, geographic information system (GIS) data (polygons and lines) can be imported. For example, any number of ecosystems or rock exposure effects can be applied to imported map polygons, resulting in visually stunning recreations and visualizations. For even more fun, WCS is designed to output animations in a variety of formats. Stereo images can also be created.

CREATING THE SHADED-RELIEF IMAGE

For the purposes of the map shown here, a rather simple shaded-relief image was created. First, 30-meter digital elevation models (DEMs) were imported (11 in all). WCS then joined the DEMs in its own native format and processed the quadrangle edges to make them as seamless as possible. The resulting digital landscape was vertically exaggerated 1.5 times to help emphasize the surficial terrain (the theme of the geologic map). Texture and color were applied to the terrain to make it look more interesting. Ambient light, light reflected from the sky and ground, was used to add light to the shadows. Quadrangle corner tics were draped on the scene. The scene was then rendered (the equivalent of taking a photograph). While WCS can import GIS data in both geographic (latitude and longitude) and Universal Transverse Mercator coordinate systems, it works with scenes only in geographic coordinates. To reduce the effects of the curved earth on the resulting image, the shadedrelief was computed or shown as if the camera were 300 kilometers above the earth. The final image was then geo-rectified to Idaho State Coordinate System coordinates in ArcGIS using the previously imported corner tics. Finally, the image was saved for layout.

LAYOUT

Adobe Illustrator 10 was used to lay out the map seen here. Like other layout programs, Illustrator handles text and imports multiple formats well. Illustrator excels at dealing with transparency or opacity of images. Taking advantage of the opacity function in Illustrator allows the user to drape a shaded image onto colored polygons. Illustrator also makes it easy to modify colors or add patterns to one or more geologic units throughout the map. For this map, opacity was set at 45 percent. Once the base map layers were registered and opacity was set, only basic map layout tasks remained to complete the map.

"Flattening" by Illustrator was required before generating the print file because of the shaded-relief geology in this map. In other words, the map layers were rasterized before plotting. This resulted in rather large (about 300MB) plot files.

FOLLOWUP INFORMATION

Although World Construction Set has a steep learning curve and the expected pitfalls that come with complex software, it is well worth exploring if you have a need to create geologic visualizations or stunning digital landscapes.

More information about World Construction Set is available online at <http://www.3dnature.com>. This site has many great examples of visualizations from a variety of disciplines, including geology. There are numerous web sites devoted to fractal geometry. I have included just two here: <http://www.ics.uci.edu/~eppstein/junkyard/fractal.html> and <http://math.rice.edu/~lanius/frac/>.

SOFTWARE CITED

Adobe Illustrator–Adobe Systems. Inc. 345 Park Ave., San Jose, CA 95110-2704 USA, (800) 833-668, <http://www/adobe.com>.

World Construction Set–3D Nature, LLC, 5740 Olde Wadworth, Suite C, Arvada, CO USA 80002, <http://www.3dnature.com>.

REFERENCES

Breckenridge, R.M. and Othberg, K.L., Surficial geologic map of the Wood River Valley area, Blaine County, Idaho: Idaho Geological Survey Surficial Geologic Map Series (in progess).

Selection of an Appropriate Base Map for a Statewide Geologic Mapping Program

By Jane E. Domier and Donald E. Luman

Illinois State Geological Survey
615 East Peabody
Champaign, Illinois 61820
Telephone: (217) 244-2513
Fax: (217) 244-2785
e-mail: domier@isgs.uiuc.edu

INTRODUCTION

In 1996, the Illinois State Geological Survey (ISGS) launched an ambitious digital mapping program to characterize geology statewide at the scale of 1:24,000 using the U.S. Geological Survey (USGS) 7.5-minute topographic quadrangle map as the fundamental mapping area. Each of the 1,071 quadrangles that comprise the state eventually will be mapped for bedrock surface, surficial geology, drift thickness and other relevant geologic factors, depending on the area. Historically, the ISGS has utilized film separates from the published versions of USGS 1: 24,000-scale topographic quadrangle maps as the primary base for geologic mapping. During the past six years, the ISGS has committed to the use of digital mapping techniques, which requires that digital base map materials be available. We review the current and future large-scale digital map products available from the USGS and identify advantages and disadvantages for their application in the ISGS Geologic Mapping Program.

USGS DIGITAL RASTER GRAPHICS (DRG)

Between 1995 and 1998, USGS produced digital raster graphics (DRGs) for some 60,000 USGS quadrangle maps of the U.S. by scanning published paper maps at 250 dots per inch (dpi) resolution. The raster image of each map was georeferenced and fit to the Universal Transverse Mercator (UTM) projection using the datum of the original, published quadrangle map. DRGs are easily imported into GIS software with no additional processing necessary. Map features are symbolized exactly as in the paper map, using standardized symbols and colors to duplicate the line and fill colors of the published map. Most DRGs have higher levels of "color noise" (multiple pixel colors within one feature color) than was originally intended, making it impossible to separate certain features by color in order to remove them (fig. 1). DRGs are available from USGS for $1.00 each, or Illinois DRGs are free for downloading from the ISGS Clearinghouse at http://www.isgs.uiuc.edu/nsdihome/webdocs/drgs/. However, the low resolution and difficulty in removing colors such as green (forested areas) and pink (urban areas) have made DRGs unacceptable as the base for geologic mapping.

Since early 1999, updated DRGs are only created when a USGS topographic quadrangle map is revised and a new edition of the map is published. Topographic maps are revised using all-digital methods by scanning copies of the map's feature separates at a high resolution, then revising these scans using a digital ortho-photograph and raster editing software. The digital feature separates are used to publish a new revised edition of the paper map. When these digital separates are combined to make a new 500 dpi DRG, they are extremely clean, without the color "noise" that is characteristic of the older DRG files scanned from paper maps.

In October 2001, a revised standard was implemented by the USGS for a new-version DRG product, with a resolution of 500 dpi and a color palette of up to 256 colors. Only about 1,000 DRGs per year are produced at this resolution. These revised standard DRGs are created when revisions are made to a quadrangle map, but there is no program or funding for systematic updating of the current DRGs to the new standard.

USGS DIGITAL LINE GRAPHS (DLG)

Over the past 20+ years, the USGS has produced digital line graph (DLG) files. DLGs are digital vector representations of topographic and planimetric map features derived from either aerial photographs or from cartographic source materials, using manual and automated digitizing methods. Large-scale (1:24,000) DLG data primarily are derived from USGS 7.5-minute topographic quadrangle maps. These data files contain selected base categories of cartographic data including:

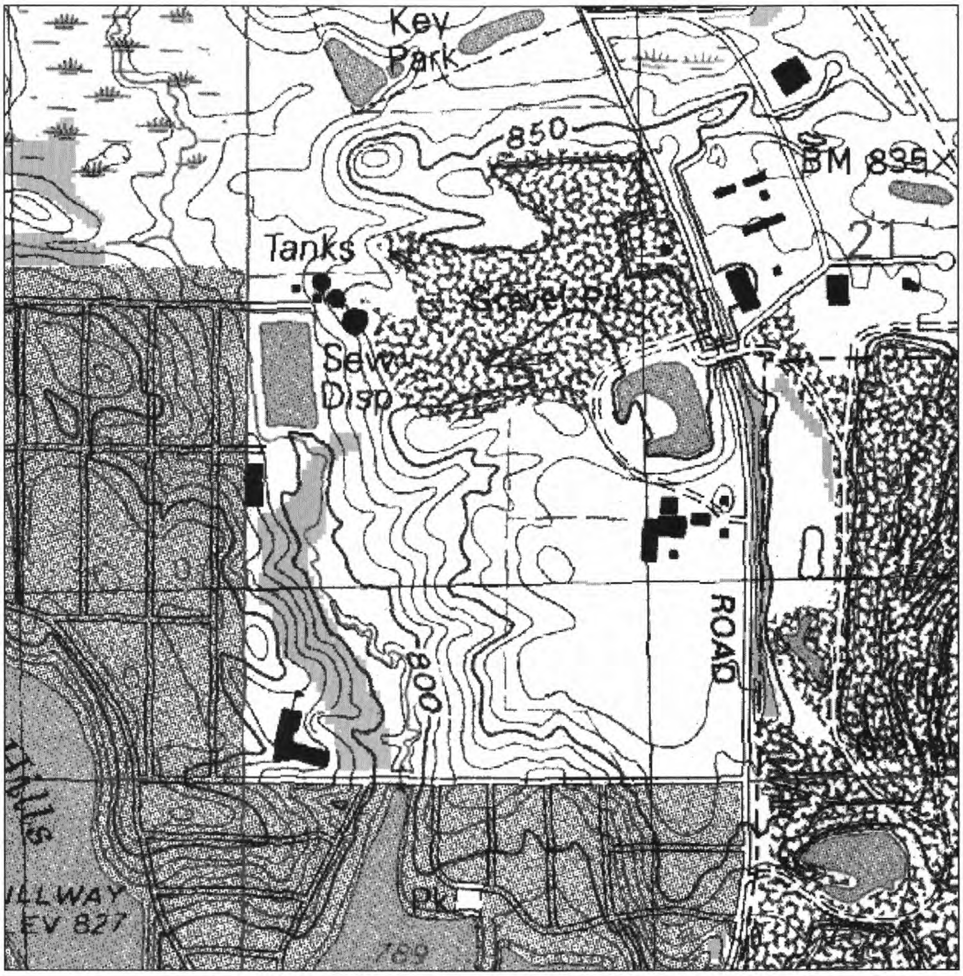


Figure 1. USGS original version Digital Raster Graphic (DRG).

1) Political boundaries and administrative boundaries
2) Hydrography
3) Public Land Survey System
4) Transportation data (includes 3 separate files: Roads and Trails, Railroads, and Miscellaneous transportation)
5) Other significant manmade structures
6) Hypsography
7) Vegetative surface cover
8) Nonvegetative surface features
9) Survey control and markers

Separate feature layers allow for selection or omission of map features, and the vector data provides high quality line work for publication of geological maps (fig. 2). A good knowledge of GIS software is required for processing DLG files. The raw files must be converted to a format compatible with GIS software, and then each feature type must be symbolized to recreate the appearance of the original topographic quadrangle map.

In the state of Illinois, the availability of DLG data files is limited (fig. 3). In fact, the majority of Illinois 7.5-minute quadrangle maps have no DLGs at all, while a minority has some or all of the DLG feature layers available. Existing DLGs are free for downloading from USGS or can be ordered for $1.00 per quadrangle. USGS prices for existing digital cartographic data are available at <http://edcsns17.cr.usgs.gov/helpdocs/prices.html>. New DLGs can be created by USGS for a much greater cost. In the past ISGS has ordered updates of four basic feature layers (hypsography, hydrography, transportation and PLSS) for approximately $7,415.00 per quadrangle. Because the DLGs do not include cartographic text, the mylar source materials for the cartographic text were also purchased, at $60.00 per mylar, (mylar map separates are no longer available from USGS, as of August 1st, 2003.) Given limited budget resources for base maps, ISGS has only been able to obtain the basic DLG feature files for 3 or 4 quadrangles per year.

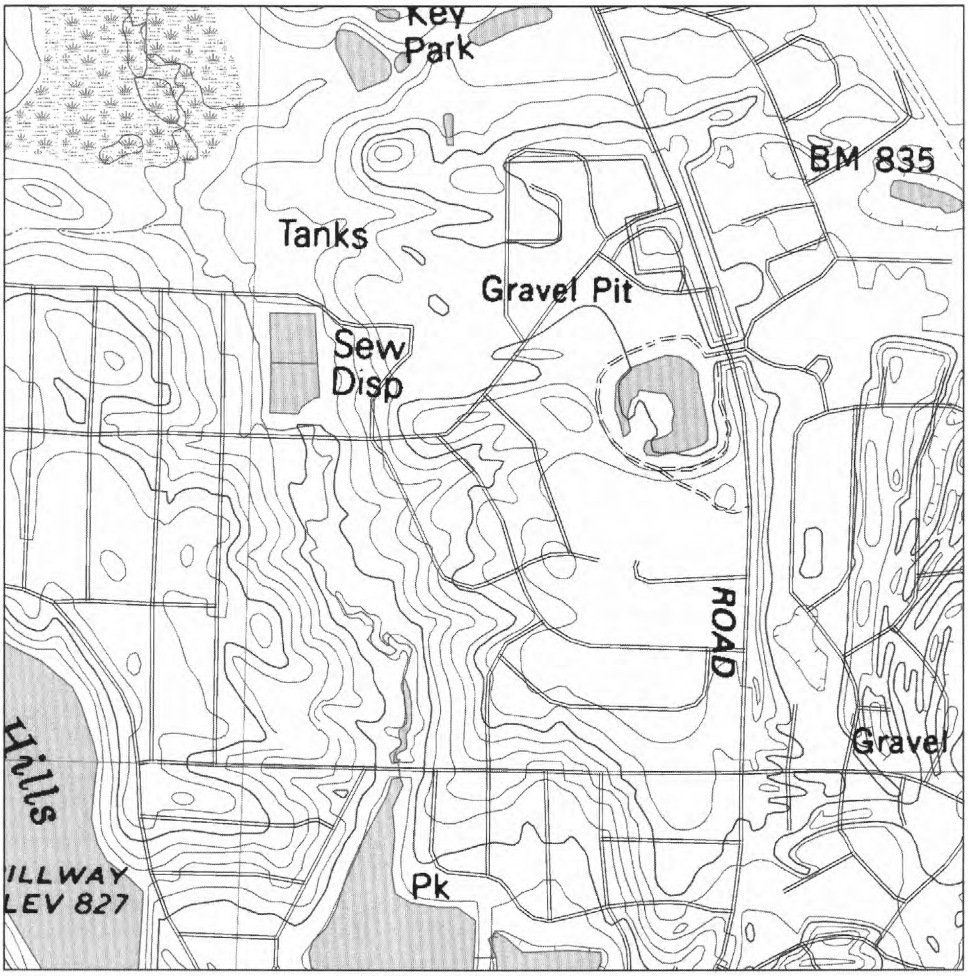


Figure 2. Digital Line Graph(DLG) with raster text scanned from mylar separate.

RASTER FEATURE SEPARATES (RFS) AND RASTER COLOR COMPOSITES (RCC)

At the time of original compilation of the 7.5-minute quadrangle maps (approximately 1945 to 1990), mylar materials were used for the production of the printed maps. All the features on one mylar separate were of the same color, but all features of the same color were not always on one separate. One quadrangle may have 5 to 20 mylar feature separates, depending on the time period it was made and the production guidelines during that time.

In the late 1980s, digital drafting tools began to replace scribing in the map revision process. Mylar separates are now scanned at 1000 dpi or greater, and the resulting raster feature separates are updated using raster revision methods. These digital raster feature separates (RFS) can be combined to make negatives for printing a new map, or to create raster color composites (RCC) which contain all of the RFS layers in one color file. These RCC files are used for the creation of 'new standard' DRG files.

Nearly all USGS feature separates are available only on mylar. Raster feature separates can be made directly from the mylar separates, without any revision. In this case the RFS and RCC files will be identical to the most current published edition of the paper map. This process is still undergoing formal standards review at the USGS, and RFS products are currently produced as a full-repay joint funding agreement with the USGS.

RFS products include both 1000 dpi and 500 dpi RFS and RCC files. Cost for one quadrangle may range between $700 and $1,200, depending on the number of map separates. These files are geo-referenced and easily used in GIS systems with no additional processing.

SUMMARY OF BASE MAP ADVANTAGES AND DISADVANTAGES

Digital Raster Graphics (DRG):

- Available for every USGS quadrangle map at a resolution of 250 dpi.
- Geo-referenced to UTM projection.

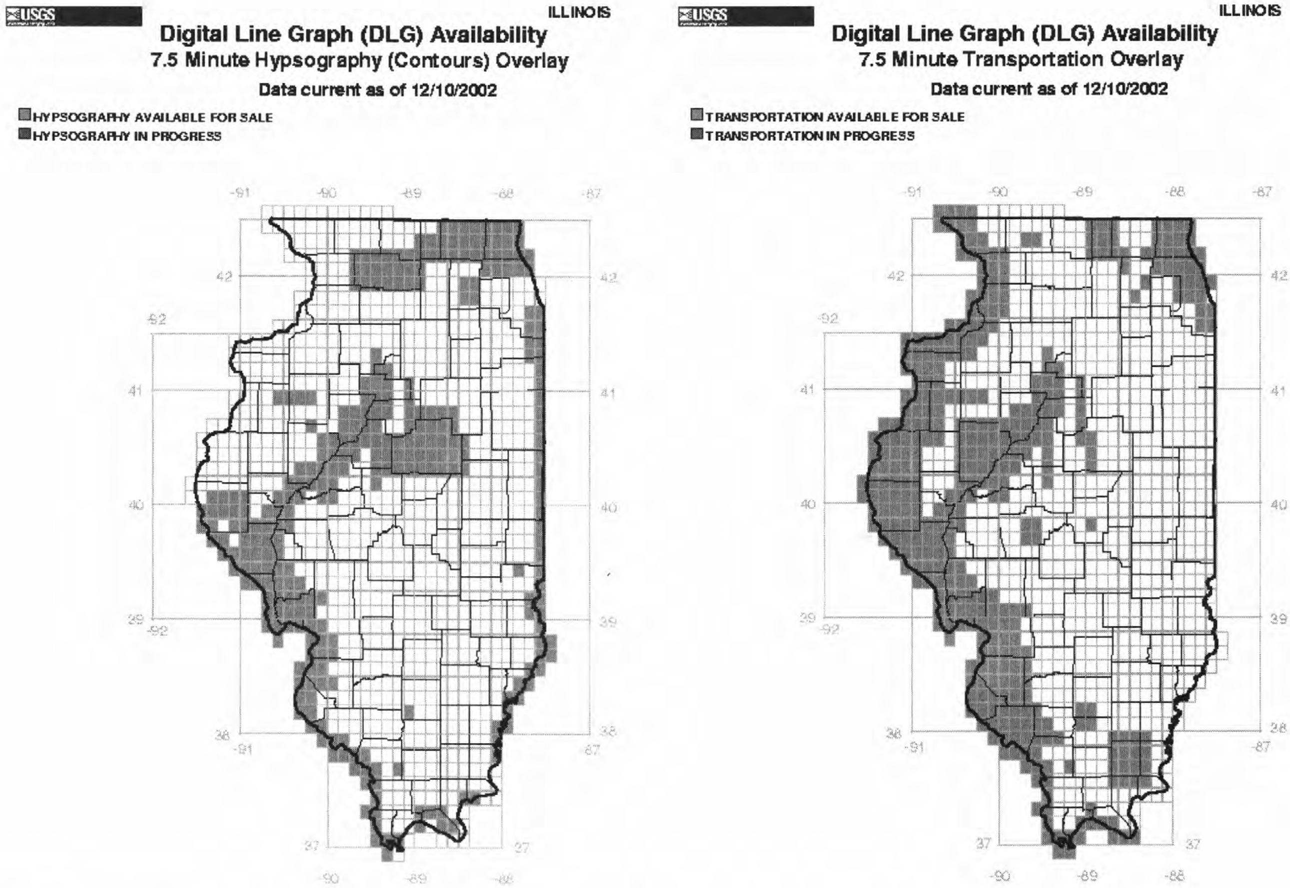


Figure 3. Availability of two USGS DLG feature layers for Illinois. Status graphics are now available at <http://geography.usgs.gov/www/products/status.html>.

- Easily imported into GIS software.
- Features are symbolized identical to paper map; no additional processing or symbolization is necessary.
- Low cost: $1.00 per quadrangle for existing DRGs from USGS. <http://edc.usgs.gov/products/map/drg.html>.
- Free for downloading through the ISGS Clearinghouse. <http://www.isgs.uiuc.edu/nsdihome/webdocs/drgs/>.
- Low quality base map for final map publication.
- Colors for features contain multiple colored pixels, making removal of built-up area, forested area, or disturbed area fills very difficult or impossible.
- A limited number of new-version DRGs are available at 500 dpi, with clean color palettes.

Digital Line Graphs (DLG)

- Vector data, producing very high quality linework for publication of geological maps.
- Vector data digitized from 7.5-minute topographic quadrangles and updates, using NAPP or NAPP-like aerial photography.
- Separate feature files allow for selection of relevant features.
- Requires good knowledge of GIS to process into base maps. DLGs must first be converted to a format compatible with GIS software, then each feature type is symbolized to recreate the appearance of the original topographic quadrangle map.
- Map text must be obtained as a raster scan of existing mylar materials.
- 1:24,000 DLGs have very limited availability in Illinois.
- Existing DLGs: download free or $1.00 per feature on CD. <http://edc.usgs.gov/products/map/dlg.html>.
- Creation /updating costs are very high ($7,415 per quad for basic 4 DLG feature layers).

USGS Raster Feature Separates (RFS) and Raster Color Composites (RCC)

- RFS files are produced from 1000 dpi resolution scans of mylar feature separates.
- Features are symbolized exactly as in the paper map; no processing or symbolization is necessary

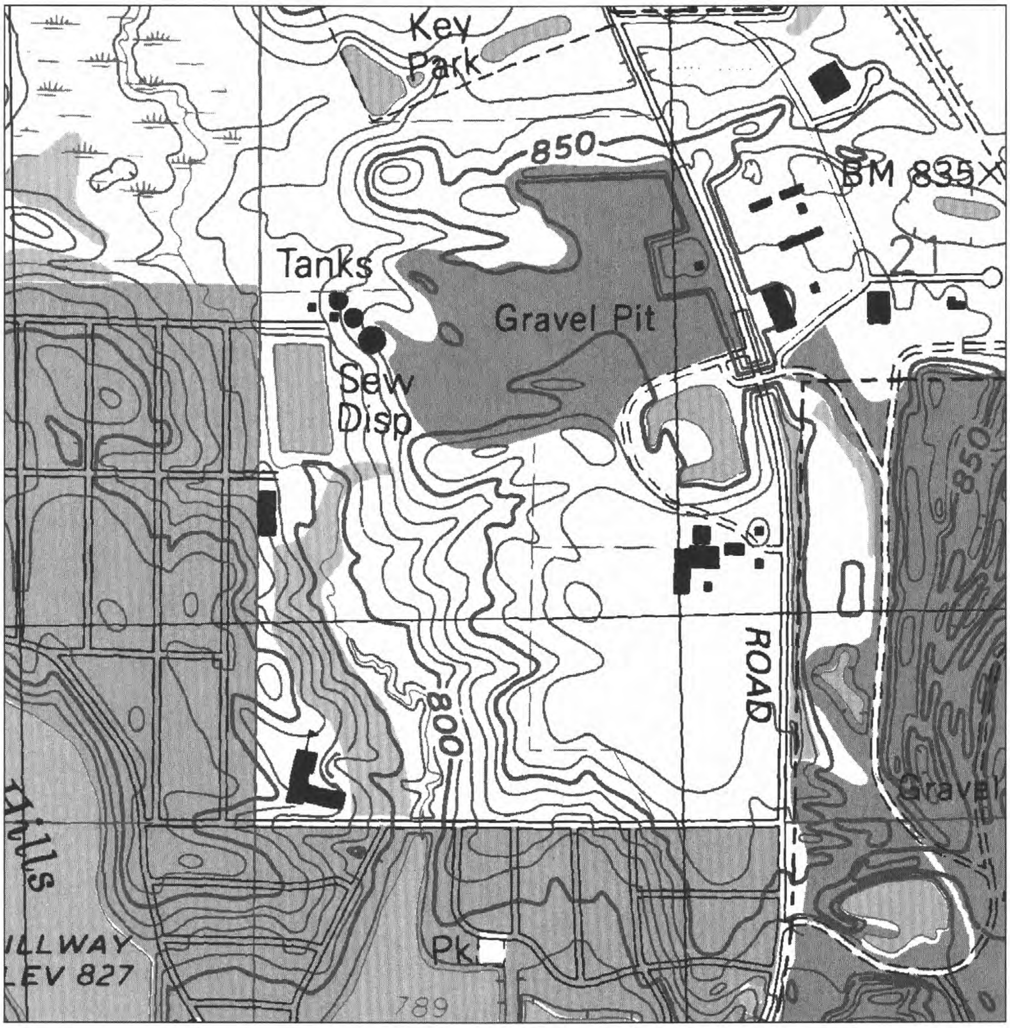


Figure 4. Raster Color Composite(RCC).

- Geo-referenced 500 and 1000 dpi RFS and RCC files are easily used in GIS systems.
- RFS and RCC files have clean color palettes, making feature separation very easy.
- Files can be requested as a full-repay cost, approximately $1000 per quadrangle.

BASE MAPS FOR THE ILLINOIS STATE GEOLOGICAL SURVEY

In 1996, at the inception of a new geologic mapping project, the geologists and GIS specialists at ISGS began developing standardized methods for producing digital geologic maps. During this time, the original-format 250 dpi USGS DRG files became available for the state of Illinois and were quickly adopted by the geologists for compiling geologic features in a GIS.

Eventually, when a map was ready for printing, the DRG was used as the base map. With only 250 dpi resolution, DRG files did not provide a sufficiently clean and clear image. Additionally, the color tints for urbanized and wooded areas could not be removed because of the combination of colors used to make each feature type. These color tint areas conflicted with the colored areas of geologic surfaces, making the map difficult to interpret.

During the same period of time, geologists at the ISGS used DLG files for the base for printing geologic maps. The original DLG files must first be converted to a format compatible with the GIS software. GIS specialists at ISGS converted the files, and created symbolization codes for dozens of feature types in the DLG layers. Symbols to match the USGS 7.5-minute topographic quadrangle maps were developed using Arc/Info.

Additionally, very few of the features in DLG files have label attributes, so that labels for roads and other features must be placed by hand. In order to avoid that time-consuming process, mylar separates of the map text were ordered from USGS, then scanned to produce a raster text layer for the map. This text usually did not include red, blue or brown text, which still needed to be hand-placed on the map.

Often, DLGs were not available for quadrangles or were in need of revision, especially in rapidly urbanizing areas where geologic mapping was prioritized. Hence, the acquisition of DLG base files was expensive and required a significant time commitment from GIS specialists and graphics personnel before the base map was ready for printing.

Early in 2002, several RFS and RCC products were created for a project in northern Illinois. After careful consideration of these raster files, ISGS decided to use these products for publication of geologic maps in 2003. The cost is much lower (approximately $1,000 per quadrangle), so it was possible to order 20 RFS base maps with annual basemap funding. The hope is that ISGS will be able to publish geologic maps more quickly with less processing time required for the RFS files, while benefiting from the high quality of the RFS basemaps.

RELATED INFORMATION

Domier, J. E., 2003, Retiring of the USGS map separates and the emergence of the USGS raster feature separates: presentation at Illinois Mapping Advisory Council meeting, November 28, 2003, Springfield, Illinois. (http://ncgmp.usgs.gov/ngmdbproject/dmt/domier03.html).

REFERENCES

USGS web site, Digital Raster Graphics (http://topomaps.usgs.gov/drg), and Digital Line Graphs, (http://edc.usgs.gov/products/map/dlg.html).

USGS Standards for Raster Feature Separates, (http://rockyweb.cr.usgs.gov/nmpstds/rfsstds.html).

Kentucky Land-Use Planning on the Web

By Gerald A. Weisenfluh and Daniel I. Carey

Kentucky Geological Survey
University of Kentucky
228 Mining and Mineral Resources Bldg.
Lexington, KY 40506
Telephone: (859) 257-5500
Fax: (859) 257-1147
e-mail: jerryw@uky.edu, carey@uky.edu

INTRODUCTION

Since 1996, the Kentucky Geological Survey (KGS) has been converting 707 7.5-minute geologic maps into digital formats. As that effort nears completion, attention is turning to a variety of uses for the newly created digital data. One of those uses is the creation of new compiled map series. A 1:100,000-scale, 30 x 60 minute map series and a 1:62,500-scale county map series have been initiated. The new maps have the benefit of regional perspective with the fidelity of 1:24,000-scale detail provided by the larger-scale source data.

The digitally formatted map data also permit creation of derivative maps that portray selective properties of the geology for special applications. For example, by relating information about the solubility of limestone units, maps of karst potential can be made for transportation or community planners. The derivative maps have an advantage over traditional geologic maps in that they are simpler to understand and they relate the geologic information with terminology that is appropriate to the user community. As in most states, land-use planning is an increasingly important issue in Kentucky, and geologic input is of critical importance. KGS is developing data related to geologic map units to support land-use assessments and is using Internet mapping services to provide these data to end users.

LAND-USE PLANNING

The target users for land-use information typically are nongeologists with a variety of specialized backgrounds. Many are members of planning councils who may have training in planning, while others are land developers or lay citizen members. The diverse backgrounds of these users make it a challenge to communicate geologic issues and information on geologic maps. Common sense dictates the preparation of simple maps that highlight information about particular issues. A good analog for this type of service is the National Resources Conservation Service (NRCS) digital soils data (SSURGO) <www.ky.nrcs.usda.gov/technical/GIS/index.html>. NRCS has developed extensive databases about the characteristics of soil units that can be used to make derivative maps for a host of problems, using the soil map units as a base.

A geologic engineering report for Fayette County, Ky. (Johnson, 1966) used a philosophy similar to the NRCS model, and KGS has used this report as a basis for developing Web services for land-use planning from the digital geologic map data.

MAP DEVELOPMENT

One of the goals in preparing the land-use maps is to simplify terminology. Most of the users are unfamiliar with the specialized technical jargon found on geologic maps. Therefore, the first step is to reclassify the map units into rock type designations that are meaningful to the user (for example, a simple classification based on lithologic characteristics). This task requires an analysis of the original map unit descriptions, review of pertinent literature, and field inspection in local areas for which the maps are being made. The second task is to determine what the most important land-use issues are for the map area (in this map series, counties). Local experts, such as the NRCS field agents or practicing geologists, are consulted to create a suitability classification for a spectrum of applications. Examples of applications are suitability for septic systems or difficulty of excavation for building foundations and basements.

INTERNET MAP SERVICE

The county land-use maps were originally formatted as print-on-demand publications. It soon became clear that delivery of the information on the Internet would decrease publication time and greatly enhance the ability to link to additional resources. A draft ArcIMS Internet map service <http://kgsweb.uky.edu/arcimsSearch.asp> has been developed that integrates the reclassified geologic map data with soils data. The map service also provides functions that enable users to zoom to an area of interest, then link to other map services (such as water well locations) available for the same area. Additional resources are provided that explain the role of geology in land-use planning.

REFERENCES

Johnson, C.G., 1966, Engineering geology of the Lexington and Fayette County, Kentucky, area: U.S. Geological Survey Open File Report, 32 p.

The National Park Service Geologic Resources Inventory "From Paper to Digital: Exploring a Geologic-GIS Map"

By Stephanie A. O'Meara[1], Trista L. Thornberry[1], Anne R. Poole[2] and Victor deWolfe[1]

[1]Colorado State University/National Park Service Cooperator
1201 Oak Ridge Drive, Suite 200
Fort Collins, CO 80525
(970) 225-3584
Stephanie_O'Meara@partner.nps.gov
Tthorn@cnr.colostate.edu
vdew@holly.colostate.edu

[2]National Park Service – Geologic Resources Division
12795 West Alameda Parkway
Lakewood, CO 80228
(303) 969-2093
Anne_Poole@nps.gov

INTRODUCTION

Bedrock and surficial geologic maps and supporting information provide the foundation for studies of ecosystems, earth history, groundwater, geomorphology, soils, and environmental hazards such as fire history, landslide and rockfall potential. Geologic maps describe the underlying physical conditions of many natural systems and are an integral component of the physical science inventories stipulated by the National Park Service (NPS) in its Natural Resources Inventory and Monitoring Guideline. The NPS has identified geographic information systems (GIS) and digital cartographic products as fundamental resource management tools. There are few geologists employed at parks, thus these tools are particularly important to the National Park Service to aid resource managers in using geologic data for park management decisions.

The NPS is developing most of its digital products in ESRI's (Environmental Systems Research Institute) ArcView GIS. ArcView 3.X interfaces effectively with other software running on the Microsoft Windows operating system. Also, integrating a variety of tools including the NPS GIS Theme Manager and Microsoft Visual Basic graphics viewer program (both are available at http://www1.nature.nps.gov/im/apps/thmmgr/index.htm), Windows Help software, and the ArcView legend editor has allowed users to display geologic map information in a digital GIS. O'Meara (2002), Connors (2002) and Fryer (2001) present additional detail of NPS guidelines for producing and presenting digital geologic maps.

BRINGING IT ALL TOGETHER

This poster details the steps entailed in transforming a paper geologic map to a user-friendly digital geologic map and database. Our ultimate goal is to have a comprehensive digital geologic-GIS map for 273 park units across the United States. This process begins with identifying existing data and getting it into a digital format. In the hopes of preserving all aspects of the original paper map, features are identified and attributed, preserving all information related to the feature (i.e. type of feature, positional accuracy and/or concealment, feature name, measurement value etc.). As well, color and symbology are matched to the extent possible using ArcView 3.3. Each map covering a portion of a park unit is then edge-matched to adjacent maps, and appended to create a compiled map of the park unit.

The compiled map includes links to graphics such as cross sections for easy viewing. GIS analysis is facilitated for our data with attributes such as unit names, feature types, measurement values, relative ages, and other feature characteristics, which are contained in the relational database. In addition to this database, a help file is linked to the GIS data to provide detailed feature descriptions. FGDC-compliant metadata for each digital map provides supporting reference and geospatial information.

Taking Stock...An Index Map

For Zion National Park our process of building a map database began with the determining the USGS 7.5 minute geographic quadrangles that included the park. We inventoried the preexisting published maps, taking note of their map scale and format (i.e. paper or digital). With the cooperation of the Utah Geological Survey, we then acquired digital data for 9 of the 11 quadrangles, mapped at 1:24,000 scale, that comprise the park (Fig. 1).

The Next Step… Identifying Map Features and Converting Data

For every map, a different set of geologic features must be captured. Features range from geologic unit areas to sample point localities. The National Park Service uses ArcView 3.X as its standard software. As mandated by this software, all features are assigned a spatial feature type such as polygon, arc, or point. In a cooperative relationship with the Utah Geological Survey (UGS), we received and converted their existing GIS digital data into our National Park Service Geology-GIS Data Model (O'Meara, 2003). Table 1 lists features present in Zion National Park, their spatial feature type or representation, and the corresponding NPS Geology-GIS Data Model data layer.

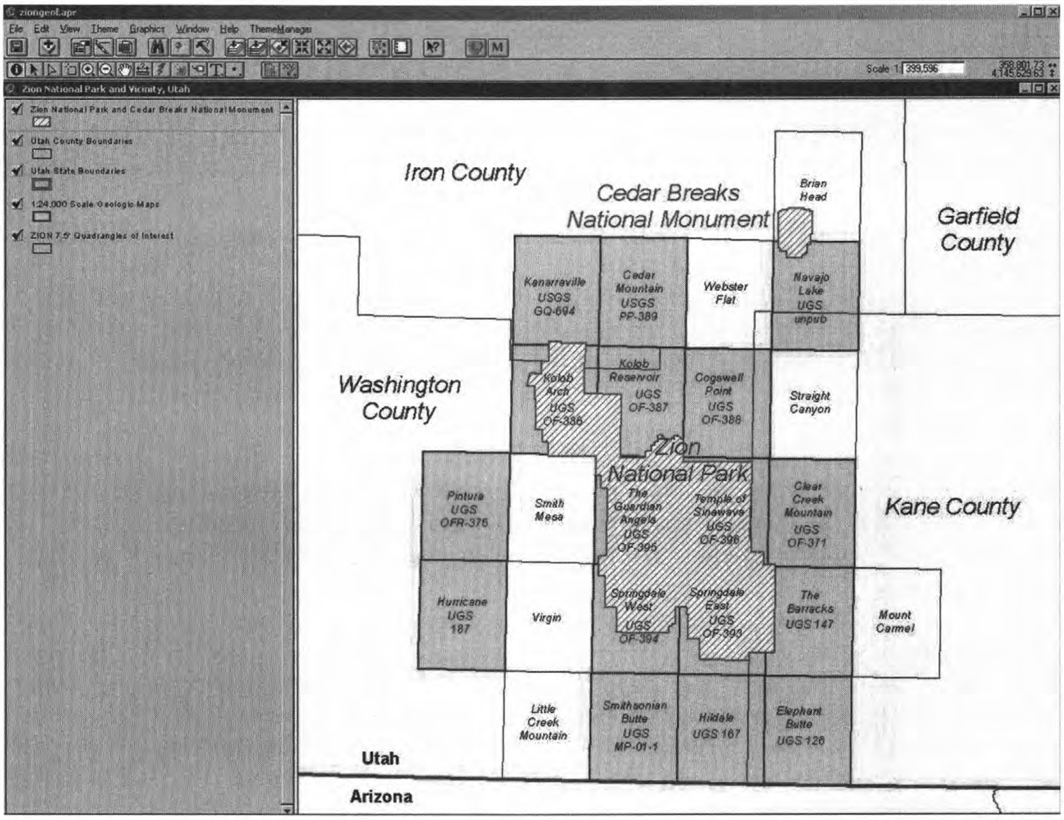


Figure 1. Index map of 1:24,000-scale geologic maps (shaded areas) for Zion National Park displayed in ArcView 3.3. Image shows park boundary extent (diagonal hatched area), USGS 7.5 minute quadrangles of park interest, Utah county boundaries, and Utah-Arizona state boundary. Nearby Cedar Breaks National Monument is also displayed.

Table 1. Identified geologic features, their spatial representation type, and their corresponding NPS Geology-GIS Data Model data layer (coverage). These apply to geologic maps for Zion National Park.

Geologic Feature Type	Spatial Feature Type	NPS Data Model Coverage
Geologic Unit Areas	Polygon	Area Geologic Units and Contacts (GLG)
Geologic Units Mapped as Linear Features	Arc	Linear Geologic Units (GLN)
Geologic Units Mapped as Point Features	Point	Point Geologic Units (GPT)
Geologic Fault Lines	Arc	Geologic Faults (FLT)
Geologic Fold Axes	Arc	Geologic Folds (FLD)
Geologic Field Attitude Measurements	Point	Geologic Attitude Observation Points (ATD)
Prospects, Adits, Shafts and Other Mine Related Features	Point	Mine and Mine Related Point Features (MIN)
Volcanic Vent Sources	Point	Volcanic Point Features (VNT)
Geologic Sample Localities	Point	Geologic Sample Points (SAM)
Geologic Structure Contour Lines	Arc	Contour and Other Lines (LN#)
Lineaments in Surficial Deposits	Arc	Contour and Other Lines (LN#)
Joints Mapped as Linear Features	Arc	Linear Joints (JLN)
Geologic Cross Section Lines	Arc	Cross Section Lines (SEC)
Landslide Scarp Lines	Arc	Linear Hazard Features (HZL)
Collapse Feature Localities	Point	Hazard Point Features (HZP)

The Barracks Quadrangle Digital Geologic Map…One of 9

Figures 2a and 2b show the converted digital map and a scan of the original published paper map for one quadrangle at Zion National Park. Our goal is to have a digital map that "looks and feels" like the published source map. For example, we sampled colors from the original map to use for the presentation of the area geologic units, and we selected and scaled symbols available in ArcView 3.3 legend palettes to best reproduce the symbology of the original map.

Many Become One…9 Quadrangles Make a Compiled Zion National Park Digital Geologic Map

Upon conversion of inherited Utah Geological Survey data, area and arc/line data layers (coverages) were edge-matched to align continuous features from one quadrangle map to another. These coverages were then appended to create a compiled coverage for the entire park map (fig. 3).

It's Hot…Linked

Using the NPS GIS Theme Manager in ArcView, a cross section graphic can be displayed at the user's whim with a simple click of the mouse (fig. 4).

Attribution, Attribution, Attribution! Its Not Just Digital Cartography Anymore

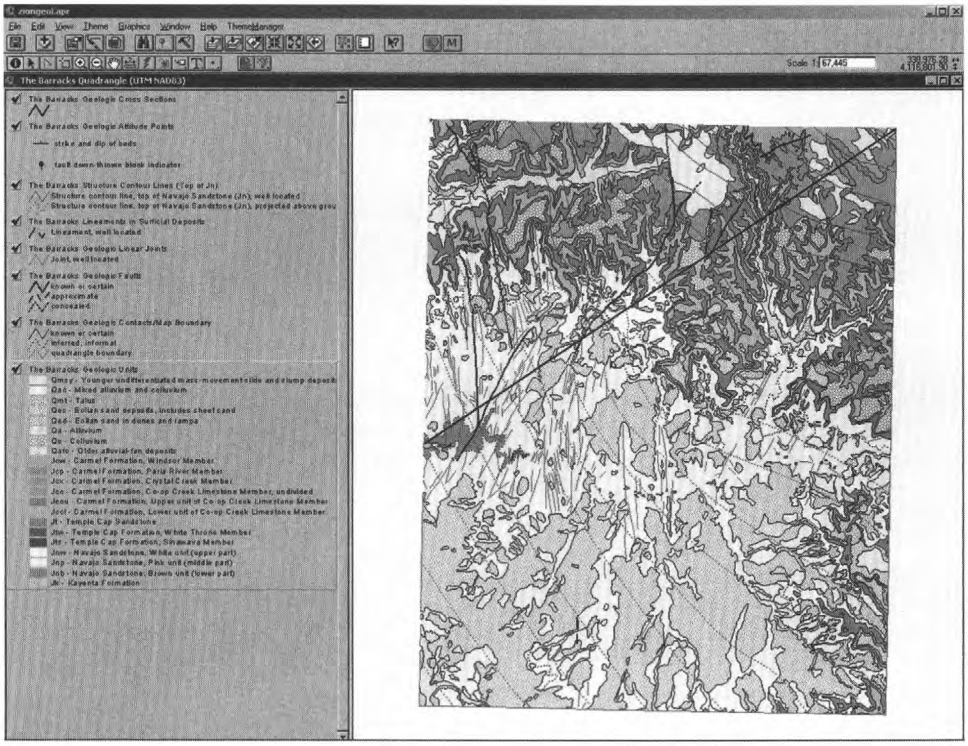
In pursuit of a useful GIS, all features captured in the digital geologic map are attributed with descriptive information contained on the paper map. Our National Park Service Geology-GIS Data Model dictates how information such as: 1) feature type, 2) feature positional accuracy and/or concealment, 3) feature name, 4) measurement value, 5) relative age, 6) feature notes, and 7) source map identification is assigned as attribution. Figure 5 shows Data Model attribution for geologic units in Zion National Park and vicinity.

The Database Enters the Picture

With relational tables storing additional feature attribution such as geologic age, formation name, member name, geologic age, and lithologic type, a database emerges. The added functionality allows for more advanced querying and geospatial analysis.

Help! I Need a Help File, Not Just Any Help File

To preserve even more data from the original source map, ancillary information such as map unit descriptions, map notes, references, figures and reports are compiled into a Windows help file (fig. 6). These help files feature



☑ The Barracks Geologic Cross Sections

☑ The Barracks Geologic Attitude Points
- strike and dip of beds
- fault down-thrown block indicator

☑ The Barracks Structure Contour Lines (Top of Jn)
- Structure contour line, top of Navajo Sandstone (Jn), well located
- Structure contour line, top of Navajo Sandstone (Jn), projected above grou

☑ The Barracks Lineaments in Surficial Deposits
- Lineament, well located

☑ The Barracks Geologic Linear Joints
- Joint, well located

☑ The Barracks Geologic Faults
- known or certain
- approximate
- concealed

☑ The Barracks Geologic Contacts/Map Boundary
- known or certain
- inferred, informal
- quadrangle boundary

☑ The Barracks Geologic Units
- Qmsy - Younger undifferentiated mass-movement slide and slump deposit
- Qac - Mixed alluvium and colluvium
- Qmt - Talus
- Qes - Eolian sand deposits, includes sheet sand
- Qed - Eolian sand in dunes and ramps
- Qa - Alluvium
- Qc - Colluvium
- Qafo - Older alluvial-fan deposits
- Jcw - Carmel Formation, Windsor Member
- Jcp - Carmel Formation, Paria River Member
- Jcx - Carmel Formation, Crystal Creek Member
- Jcc - Carmel Formation, Co-op Creek Limestone Member, undivided
- Jccu - Carmel Formation, Upper unit of Co-op Creek Limestone Member
- Jccl - Carmel Formation, Lower unit of Co-op Creek Limestone Member
- Jt - Temple Cap Sandstone
- Jtw - Temple Cap Formation, White Throne Member
- Jts - Temple Cap Formation, Sinawava Member
- Jnw - Navajo Sandstone, White unit (upper part)
- Jnp - Navajo Sandstone, Pink unit (middle part)
- Jnb - Navajo Sandstone, Brown unit (lower part)
- Jk - Kayenta Formation

Figure 2a. The Barracks digital geologic map displayed in ArcView 3.3.

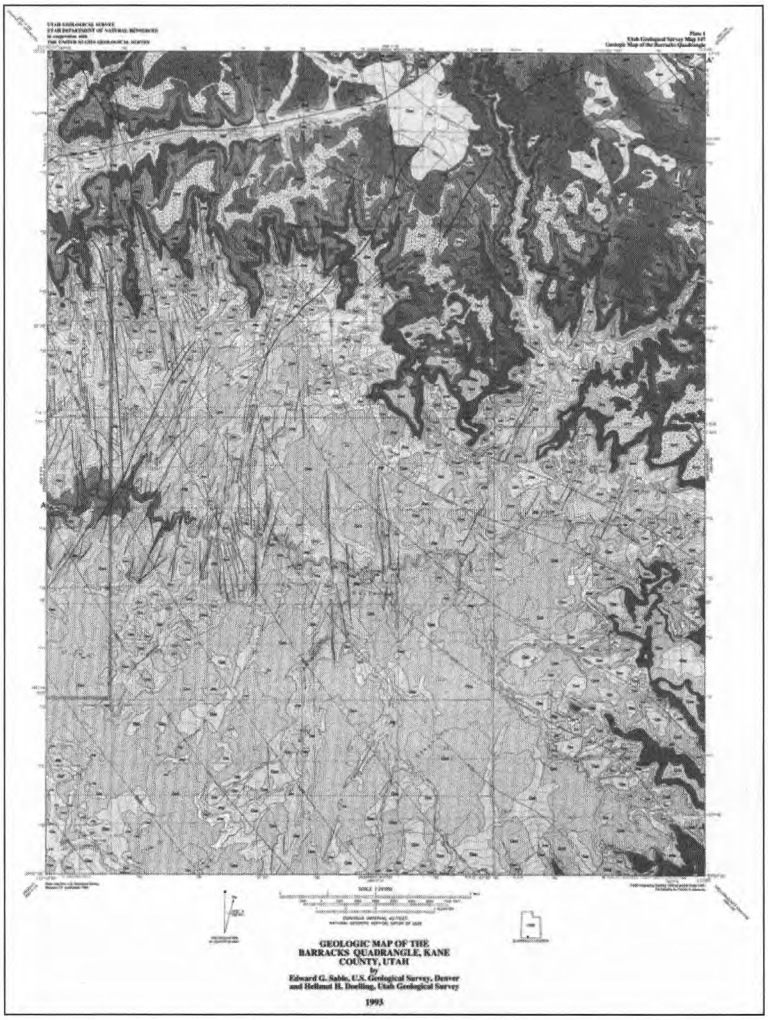


Figure 2b. For comparison, a scanned image of The Barracks paper source map.

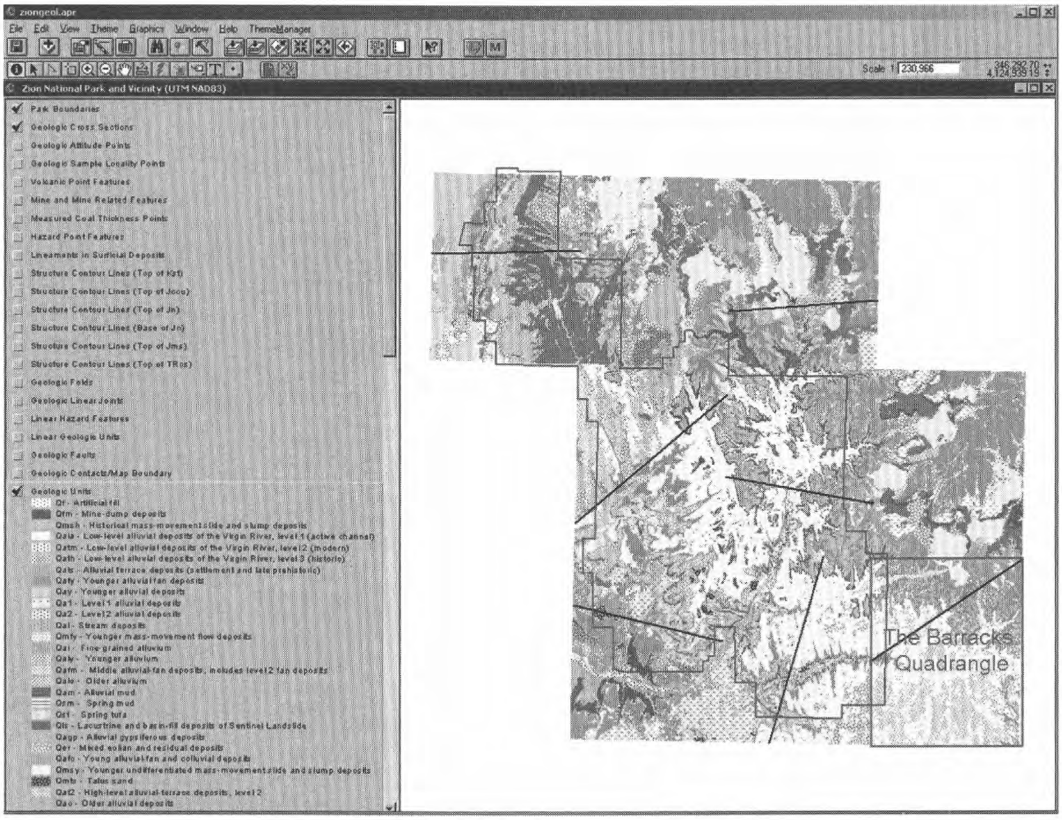


Figure 3. The compiled Zion National Park and vicinity digital geologic map, displayed in ArcView 3.3.

keyword searchability and topical organization. The help file is linked to the area geologic units of the digital geologic map.

Those Two Four Letter Words…Metadata, FGDC Style

In compliance with Federal Geospatial Data Committee guidelines, metadata is produced for all GIS data layers (fig 7). To produce FGDC metadata records, metadata template files are populated with information specific to the GIS data including source map reference information, keywords and process steps used in the creation of the data. The edited template file is then imported into ArcCatalog, which harvests map data layer-specific geospatial information such as projection, types of features, and spatial domain.

CONCLUSION

The National Park Service has identified digital geologic-GIS maps as an integral component in managing the natural resources of its parks. The process of producing a digital geologic-GIS map begins with identifying existing data. If digital data is available, the next step is getting the data into a standard format using the NPS Geology-GIS Data Model. The Data Model dictates how geologic features are captured and organized in a digital GIS, and how information associated with a geologic feature is recorded in GIS attribution.

For park maps that are composed of more than one map, adjacent maps are edge-matched to align features that are continuous from one map to another. These individual maps are then appended to create one park map. Ancillary information such as geologic unit descriptions and map notes as well as cross section graphics associated with source maps are converted into digital format, and presented in a digital GIS using tools developed by the NPS. Lastly, FGDC compliant metadata is produced for each geologic data layer.

SOFTWARE REFERENCES

ArcInfo 8.3, ArcView 3.3 - Environmental Systems Research Institute (ESRI) Inc., 380 New York St., Redlands, CA 92373, http://www.esri.com

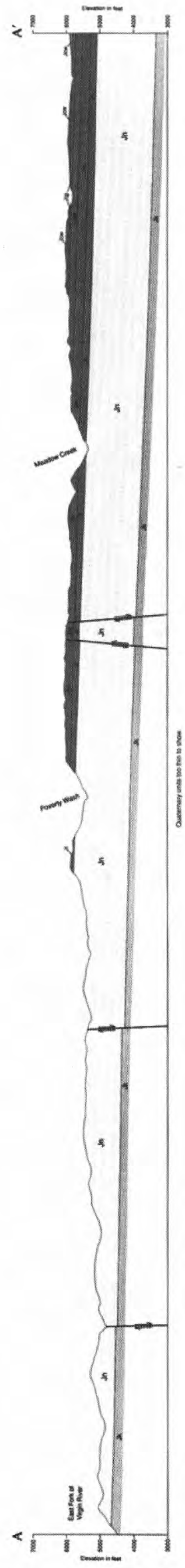


Figure 4. A geologic cross section, displayed in ArcView 3.3 using the NPS GIS Theme Manager extension.

REFERENCES

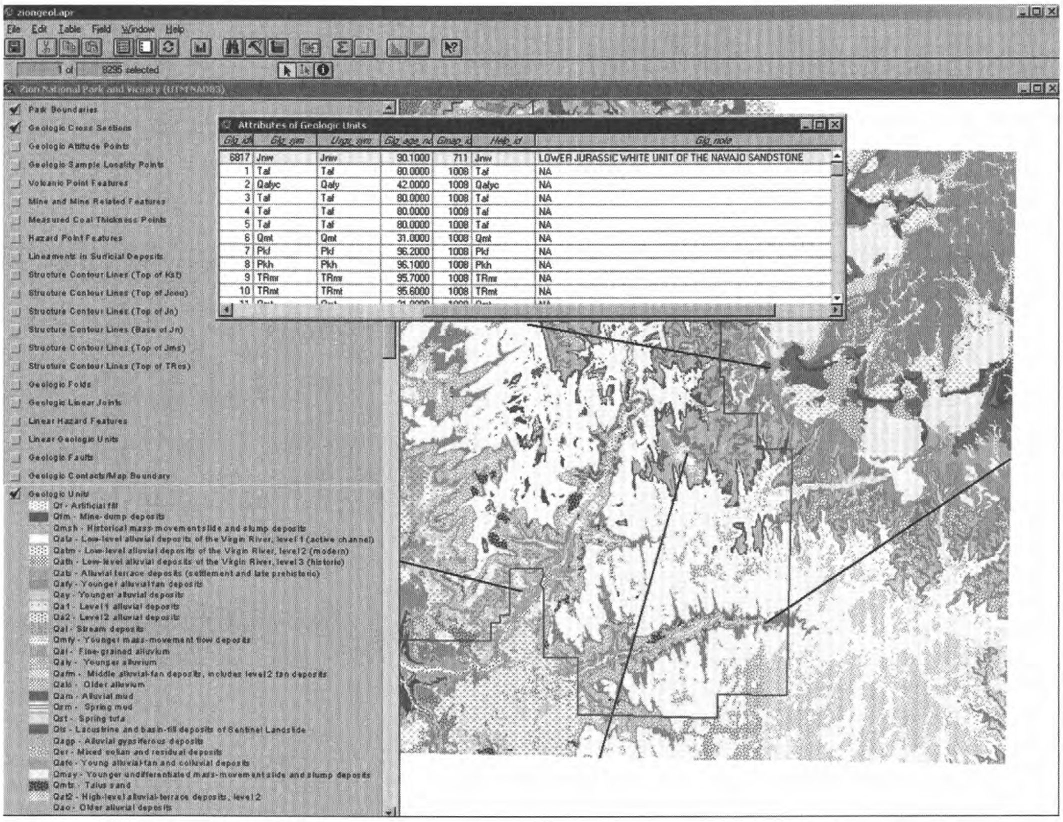
Connors, T., O'Meara, S. A., Thornberry, T. L., Gregson, J. D., Poole, A. R. and Heise, B., 2002, The National Park Service Geologic Resources Inventory: An Update, *in* D.R. Soller, ed. Digital Mapping Techniques '02—Workshop Proceedings: U.S. Geological Survey Open-File Report OF-02-370, <http://pubs.usgs.gov/of/2002/of02-370/connors.html>.

Fryer, S., Gregson, J., Connors, T., Poole, A., and Heise, B., 2001, The National Park Service Digital Geologic Map Model: Transformation from Paper to Digital, Featuring Legends, Cross Sections, Map Notes and Keyword Searchability, *in* D.R. Soller, ed. Digital Mapping Techniques '01—Workshop Proceedings: U.S. Geological Survey Open-File Report 01-223, p.87-100, <http://pubs.usgs.gov/openfile/of01-223/fryer.html>.

O'Meara, S. A., Poole, A. R., deWolfe, Victor and Thornberry, T. L., 2003, National Park Service Geology-GIS Data Model, <http://www1.nature.nps.gov/im/gis/GeologyGISDataModel.htm>.

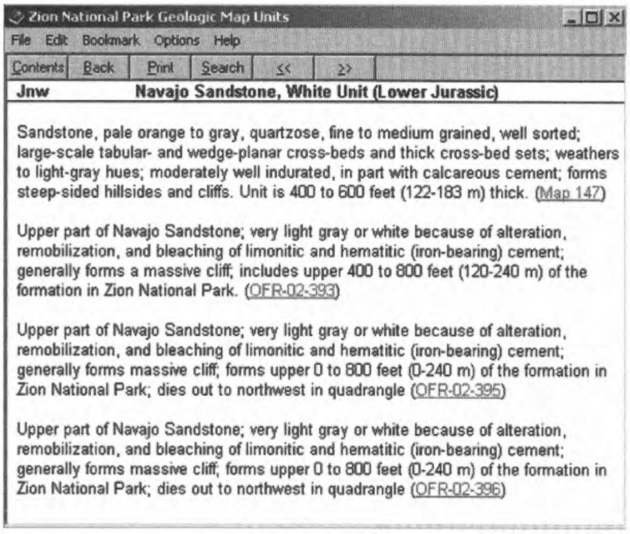
O'Meara, S. A., Connors, Tim, Thornberry, T. L., Gregson, J. D., Poole, A. R., and Heise, B., 2002, The National Park Service Geologic Resources Inventory and from Paper to Digital: a Geologic Map's Odyssey, *in* D.R. Soller, ed. Digital Mapping Techniques '02—Workshop Proceedings: U.S. Geological Survey Open-File Report OF-02-370, <http://pubs.usgs.gov/openfile/of02-370/omeara.html>.

Sable, Edward G., and Doelling, Helmut H., 1993, Geologic map of the Barracks quadrangle, Kane County, Utah, Utah Geological Survey, Map 147, 1:24,000 scale.



Glg_id	Glg_sym	Usgs_sym	Glg_age_no	Gmap_id	Help_id	Glg_note
6817	Jnw	Jnw	90.1000	711	Jnw	LOWER JURASSIC WHITE UNIT OF THE NAVAJO SANDSTONE
1	Taf	Taf	80.0000	1008	Taf	NA
2	Qafyc	Qafy	42.0000	1008	Qafyc	NA
3	Taf	Taf	80.0000	1008	Taf	NA
4	Taf	Taf	80.0000	1008	Taf	NA
5	Taf	Taf	80.0000	1008	Taf	NA
6	Qmt	Qmt	31.0000	1008	Qmt	NA
7	Pkf	Pkf	96.2000	1008	Pkf	NA
8	Pkh	Pkh	96.1000	1008	Pkh	NA
9	TRmr	TRmr	95.7000	1008	TRmr	NA
10	TRmt	TRmt	95.6000	1008	TRmt	NA

Figure 5. Part of the NPS Geology-GIS Data Model attribute table for area geologic units displayed in ArcView 3.3



Jnw **Navajo Sandstone, White Unit (Lower Jurassic)**

Sandstone, pale orange to gray, quartzose, fine to medium grained, well sorted; large-scale tabular- and wedge-planar cross-beds and thick cross-bed sets; weathers to light-gray hues; moderately well indurated, in part with calcareous cement; forms steep-sided hillsides and cliffs. Unit is 400 to 600 feet (122-183 m) thick. (Map 147)

Upper part of Navajo Sandstone; very light gray or white because of alteration, remobilization, and bleaching of limonitic and hematitic (iron-bearing) cement; generally forms a massive cliff; includes upper 400 to 800 feet (120-240 m) of the formation in Zion National Park. (OFR-02-393)

Upper part of Navajo Sandstone; very light gray or white because of alteration, remobilization, and bleaching of limonitic and hematitic (iron-bearing) cement; generally forms massive cliff; forms upper 0 to 800 feet (0-240 m) of the formation in Zion National Park; dies out to northwest in quadrangle (OFR-02-395)

Upper part of Navajo Sandstone; very light gray or white because of alteration, remobilization, and bleaching of limonitic and hematitic (iron-bearing) cement; generally forms massive cliff; forms upper 0 to 800 feet (0-240 m) of the formation in Zion National Park; dies out to northwest in quadrangle (OFR-02-396)

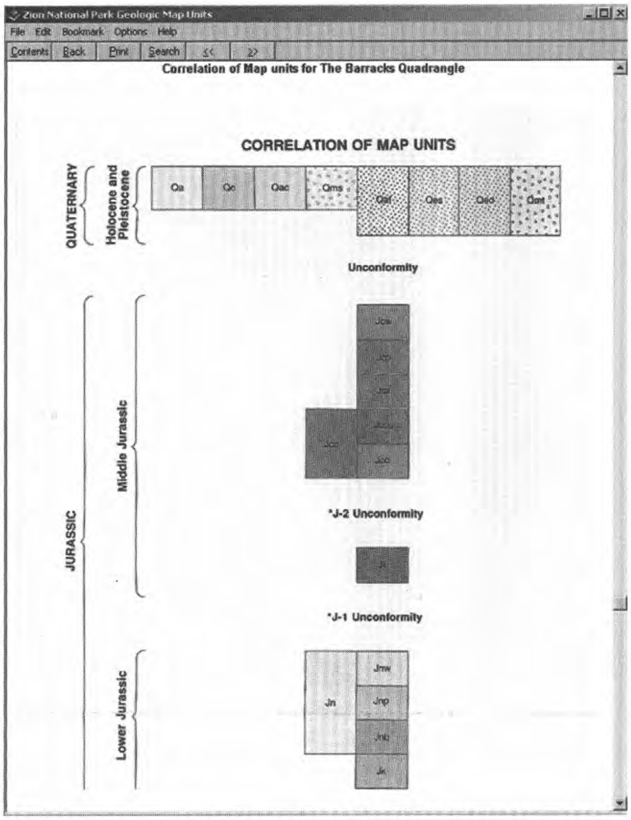
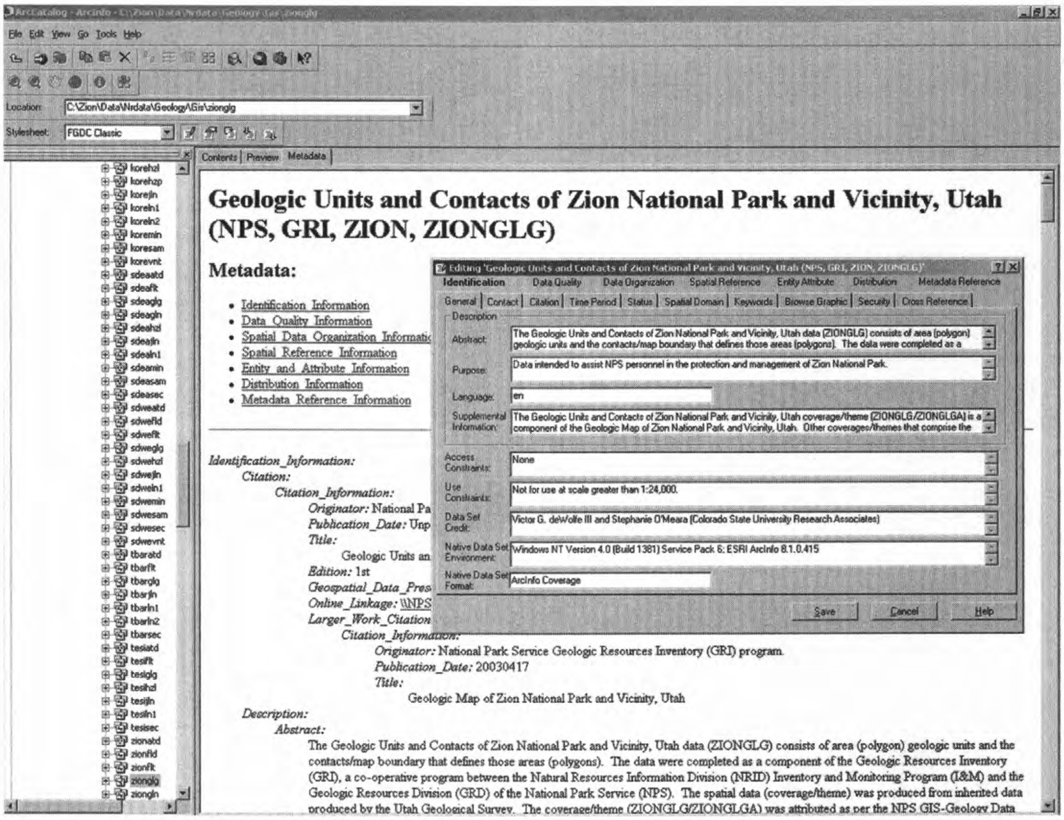


Figure 6. Title page, geologic unit description page, and correlation of map unit help file page in the Zion National Park map windows help file.



Location: C:\Zion\Data\Nrdata\Geology\Gis\zionglg

Stylesheet: FGDC Classic

Geologic Units and Contacts of Zion National Park and Vicinity, Utah (NPS, GRI, ZION, ZIONGLG)

Metadata:

- Identification Information
- Data Quality Information
- Spatial Data Organization Informatic
- Spatial Reference Information
- Entity and Attribute Information
- Distribution Information
- Metadata Reference Information

Identification_Information:
Citation:
Citation_Information:
Originator: National Pa
Publication_Date: Unp
Title:
Geologic Units an
Edition: 1st
Geospatial_Data_Pres
Online_Linkage: \\NPS
Larger_Work_Citation
Citation_Information:
Originator: National Park Service Geologic Resources Inventory (GRI) program.
Publication_Date: 20030417
Title:
Geologic Map of Zion National Park and Vicinity, Utah
Description:
Abstract:
The Geologic Units and Contacts of Zion National Park and Vicinity, Utah data (ZIONGLG) consists of area (polygon) geologic units and the contacts/map boundary that defines those areas (polygons). The data were completed as a component of the Geologic Resources Inventory (GRI), a co-operative program between the Natural Resources Information Division (NRID) Inventory and Monitoring Program (I&M) and the Geologic Resources Division (GRD) of the National Park Service (NPS). The spatial data (coverage/theme) was produced from inherited data produced by the Utah Geological Survey. The coverage/theme (ZIONGLG/ZIONGLGA) was attributed as per the NPS GIS-Geology Data

Editing 'Geologic Units and Contacts of Zion National Park and Vicinity, Utah (NPS, GRI, ZION, ZIONGLG)'

Abstract: The Geologic Units and Contacts of Zion National Park and Vicinity, Utah data (ZIONGLG) consists of area (polygon) geologic units and the contacts/map boundary that defines those areas (polygons). The data were completed as a

Purpose: Data intended to assist NPS personnel in the protection and management of Zion National Park.

Language: en

Supplemental Information: The Geologic Units and Contacts of Zion National Park and Vicinity, Utah coverage/theme (ZIONGLG/ZIONGLGA) is a component of the Geologic Map of Zion National Park and Vicinity, Utah. Other coverages/themes that comprise the

Access Constraints: None

Use Constraints: Not for use at scale greater than 1:24,000.

Data Set Credit: Victor G. deWolfe III and Stephanie O'Meara (Colorado State University Research Associates)

Native Data Set Environment: Windows NT Version 4.0 (Build 1381) Service Pack 6; ESRI ArcInfo 8.1.0.415

Native Data Set Format: ArcInfo Coverage

Figure 7. Editable FGDC compliant metadata displayed in ArcCatalog.

Meeting Constituent Needs For 1:24,000 Scale Geoscience Data In New Jersey

By Ronald Pristas

New Jersey Geological Survey
P.O. Box 427
29 Arctic Parkway
Trenton, NJ 08625-0427
Telephone: (609) 292-2576
Fax: (609) 633-1004
Email: ron.pristas@dep.state.nj.us

The New Jersey Geological Survey is distributing 1:24,000-scale geoscience data in three ways: traditional sales of hard copy maps through a Maps and Publications Sales Office, download of Open-File Map (OFM) and Geologic Map Series (GMS) maps from the Survey's internet web page, and download of geographic information system-compatible map data sets. The three distribution methods the Survey uses meets our constituents' needs by providing both hard copy and digital geoscience information.

At this time we have thirteen OFM and GMS maps available for download at <http://www.nj.gov/dep/njgs/pricelst/njgsmaps.htm>. The displayed maps (figures 1 and 2) are examples of the Open-File maps available for download as Adobe Acrobat portable document format (PDF) files.

The GIS data sets of these two quadrangles are not yet available for download but may be obtained through a written request to the New Jersey State Geologist (karl.muessig@dep.state.nj.us). Several 1:24,0000-scale county data sets are available at <http://www.nj.gov/dep/njgs/geodata/archive.htm>.

REFERENCES

Monteverde, D. H., 2000, Bedrock geology of the Roselle quadrangle, Union and Essex Counties, New Jersey: New Jersey Geological Survey Open-File Map OFM 34, scale 1:24,000, PDF file, 2.0 MB.

Stanford, S. D., 2002, Surficial geology of the Tranquility quadrangle, Warren, Sussex, and Morris Counties, New Jersey: New Jersey Geological Survey Open-File Map OFM 51, scale 1:24,000, PDF file, 3.9 MB.

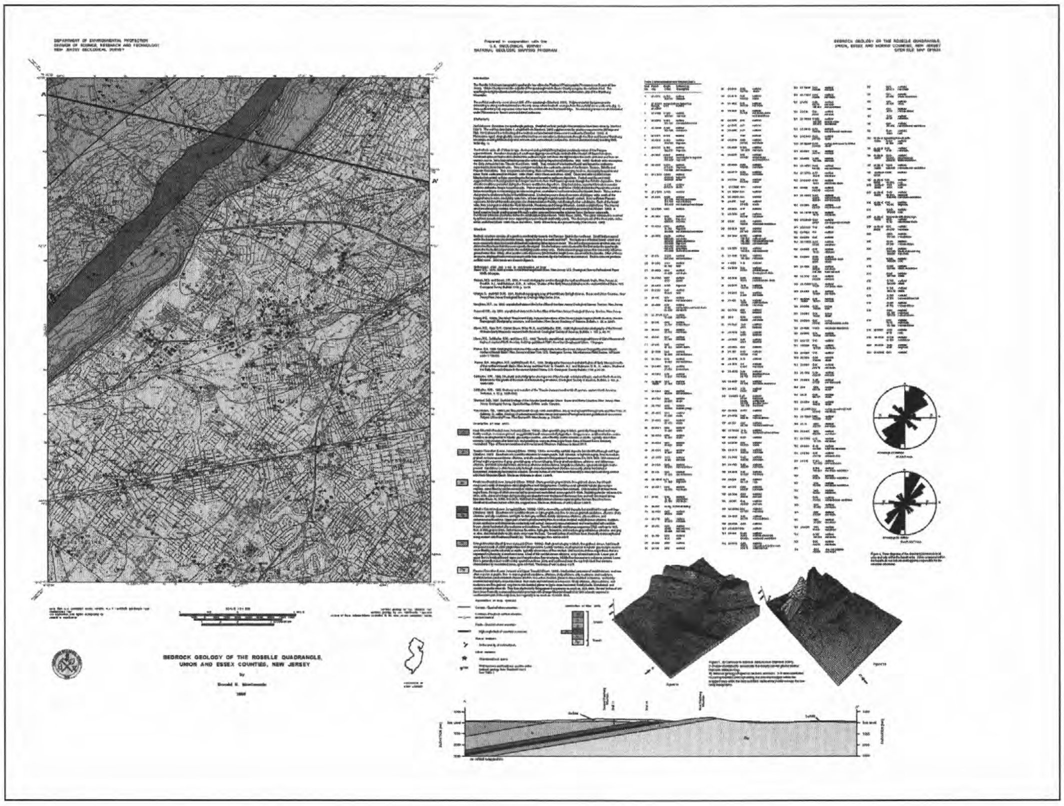
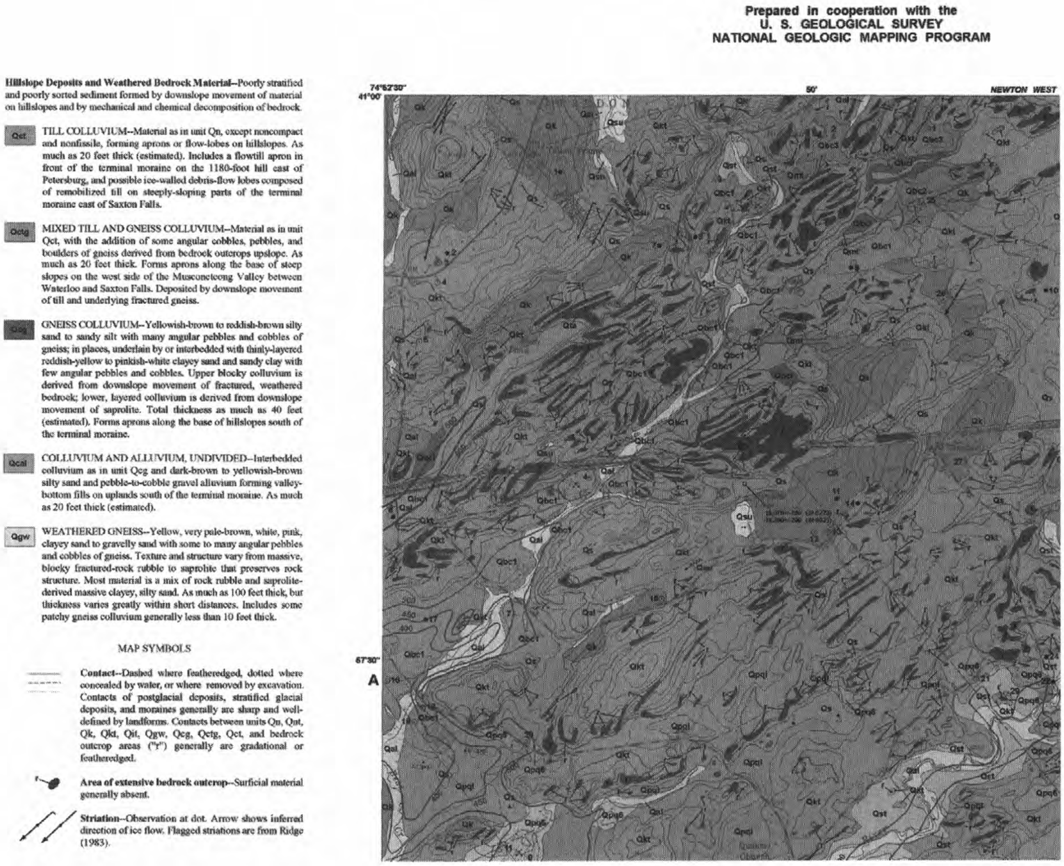


Figure 1. Black and white copy of full-color map of the bedrock geology of the Roselle quadrangle (Monteverde, 2000).



Hillslope Deposits and Weathered Bedrock Material--Poorly stratified and poorly sorted sediment formed by downslope movement of material on hillslopes and by mechanical and chemical decomposition of bedrock.

Qct TILL COLLUVIUM--Material as in unit Qn, except noncompact and nonfissile, forming aprons or flow-lobes on hillslopes. As much as 20 feet thick (estimated). Includes a flowtill apron in front of the terminal moraine on the 1180-foot hill east of Petersburg, and possible ice-walled debris-flow lobes composed of remobilized till on steeply-sloping parts of the terminal moraine east of Saxton Falls.

Qctg MIXED TILL AND GNEISS COLLUVIUM--Material as in unit Qct, with the addition of some angular cobbles, pebbles, and boulders of gneiss derived from bedrock outcrops upslope. As much as 20 feet thick. Forms aprons along the base of steep slopes on the west side of the Musconetcong Valley between Waterloo and Saxton Falls. Deposited by downslope movement of till and underlying fractured gneiss.

Qcg GNEISS COLLUVIUM--Yellowish-brown to reddish-brown silty sand to sandy silt with many angular pebbles and cobbles of gneiss; in places, underlain by or interbedded with thinly-layered reddish-yellow to pinkish-white clayey sand and sandy clay with few angular pebbles and cobbles. Upper blocky colluvium is derived from downslope movement of fractured, weathered bedrock; lower, layered colluvium is derived from downslope movement of saprolite. Total thickness as much as 40 feet (estimated). Forms aprons along the base of hillslopes south of the terminal moraine.

Qcal COLLUVIUM AND ALLUVIUM, UNDIVIDED--Interbedded colluvium as in unit Qcg and dark-brown to yellowish-brown silty sand and pebble-to-cobble gravel alluvium forming valley-bottom fills on uplands south of the terminal moraine. As much as 20 feet thick (estimated).

Qgw WEATHERED GNEISS--Yellow, very pale-brown, white, pink, clayey sand to gravelly sand with some to many angular pebbles and cobbles of gneiss. Texture and structure vary from massive, blocky fractured-rock rubble to saprolite that preserves rock structure. Most material is a mix of rock rubble and saprolite-derived massive clayey, silty sand. As much as 100 feet thick, but thickness varies greatly within short distances. Includes some patchy gneiss colluvium generally less than 10 feet thick.

MAP SYMBOLS

Contact--Dashed where featheredged, dotted where concealed by water, or where removed by excavation. Contacts of postglacial deposits, stratified glacial deposits, and moraines generally are sharp and well-defined by landforms. Contacts between units Qn, Qnt, Qk, Qkt, Qit, Qgw, Qcg, Qctg, Qct, and bedrock outcrop areas ("r") generally are gradational or featheredged.

Area of extensive bedrock outcrop--Surficial material generally absent.

Striation--Observation at dot. Arrow shows inferred direction of ice flow. Flagged striations are from Ridge (1983).

Figure 2. Black and white copy of part of the full-color map of the surficial geology of the Tranquility quadrangle (Stanford, 2002).

Problems and Solutions in the Digital Compilation and Production of the "Map of Surficial Deposits and Materials in the Eastern and Central United States (East of 102° West Longitude)"

By Charles A. Bush, Diane E. Lane, David S. Fullerton, and Nancy Shock

U.S. Geological Survey
Box 25046, Denver Federal Center, MS980
Denver, CO 80205
Telephone: (303)236–4723
Fax: (303)236–0214
e-mail: cbush@usgs.gov

ABSTRACT

The "Map of Surficial Deposits and Materials in the Eastern and Central United States (East of 102° West Longitude)" (Fullerton and others, 2003) depicts the areal distribution of surficial geologic deposits and other materials that accumulated or formed during the past 2+ million years, the period that includes all activities of the human species. These materials are at the surface of the earth. They make up the "ground" on which we walk, the "dirt" in which we dig foundations, and the "soil" in which we grow crops. Most of our human activity is related in one way or another to these surface materials that are referred to collectively by many geologists as "regolith," the mantle of fragmented and generally unconsolidated material that overlies the bedrock foundation of the continent. The map is based on 31 published maps in the U.S. Geological Survey's Quaternary Geologic Atlas of the United States (U.S. Geological Survey Miscellaneous Investigations Series I–1420). It was compiled at a scale of 1:1,000,000, to be viewed as a digital map at a nominal scale of 1:2,000,000 and to be printed as a conventional paper map at 1:2,500,000.

The map unit descriptions provide information about genesis (processes of origin) or environments of deposition; age; properties, that is, the physical, chemical, and mechanical or engineering characteristics of the materials; and thickness or depth to underlying deposits or materials or to bedrock. The map and associated database provide information about areal distribution of more than 150 types of materials. The map and database also show the maximum limits of glacial advance during selected time periods. The database is available as ArcInfo export files and ArcView shapefiles at <http://pubs.usgs.gov/imap/i-2789/>.

Preparation of the digital database consisted of the following steps:

1. The number of map units on this map is much smaller than the total number of map units on the 4° x 6° quadrangles on which the map is based. The individual map unit descriptions were cut from each published map, and each description was labeled with respect to genetic class, age or age range, and quadrangle name. These descriptions were sorted by genetic class (for example, eolian deposits, alluvium, solution residuum). Then, within the genetic classes, they were sorted by age or age class (for example, Holocene). Within each genetic or age class, the descriptions then were grouped by particle size or texture; lithology or composition; engineering properties; stratigraphic relationships; and other information in the unit descriptions. Each group of individual unit descriptions constituted a single unit on the new map. Each group was taped onto pages in notebooks and assigned a new unit name and letter symbol. The letter symbols chosen were arbitrary. The list of map units for the surficial geologic map then was prepared from the hierarchy of units organized in the notebooks. The number of map units was greatly reduced. As an example, map unit **cl** on the surficial geologic map represents 13 different map units in 11 individual 4° x 6° quadrangles. All 13 units were colluvium derived from clastic rocks (conglomerate, sandstone, quartzitic sandstone, siltstone, shale) in various combinations. The distinctions on the source maps at 1:1,000,000 are not warranted on this map. The unit descriptions for the surficial geologic map were compiled from all of the cut-and-taped descriptions that were assembled into a new map unit. The published

descriptions were generalized and simplified.
2. Each 1:1,000,000-scale, 4° x 6° quadrangle map in the Quaternary Geologic Atlas of the United States was simplified; for example, some small or narrow units were deleted or units were combined. In some quadrangles, map units were revised or modified to accommodate information that was not available when the maps were published (for example, the age of a deposit subsequently may have been revised). The contacts of the new simplified or revised map units were inked on a paper copy of each 4° x 6° quadrangle, and letter symbols were assigned to the new map units from the new generalized and simplified map descriptions.
3. Because the quadrangles of the Quaternary Geologic Atlas of the United States were compiled and printed in different projections and on different bases, the projections and bases had to be converted to a common one for publication. But first, the source geology had to be recompiled to match the digital base on which the map was to be printed, the Streams and Waterbodies GIS file from the National Atlas of the United States. This file was converted to the projection of each published Quaternary Geologic Atlas map, clipped to the area of each 4° x 6° quadrangle, and printed in blue ink on mylar at 1:1,000,000.
4. The mylar hydrographic base for each quadrangle was placed over the inked paper map on a light table. The map units then were traced in black ink on the mylar overlay. The contacts of surficial deposits and materials in and adjacent to major valleys were traced onto the overlay by matching stream junctions, river bends, lakes, reservoirs, and other components of the hydrography on the paper map and the overlay. That procedure was accomplished in increments. Within each major valley, geologic contacts were "fitted" on the mylar within an area approximately 1 inch square, then "fitted" in an adjacent square. When the geology in and adjacent to all of the major valleys had been transferred to the mylar base, the geology in the areas between valleys was "fitted" in increments by using the valley deposits and materials, hydrographic features (for example, minor streams and lakes), state boundaries, and other guides.
5. The limits of selected glacial advances indicated on the Quaternary Geologic Atlas maps also were transferred to the mylar overlays. Those lines were used for most of the glacial limits coverage, but were altered in some areas to incorporate more recent geologic mapping.
6. A paper copy of the completed mylar quadrangle map was produced. Letter symbols were added to the paper map, and the map units were differentiated by color (using colored pencils). The colored paper sheet with letter symbols served as a guide for attribution of the map units.
7. The mylar plots of the geology then were scanned, vectorized with the LT4X computer program, and converted into ArcInfo (ESRI) coverages. Polygons and lines then were attributed. Map data for the individual quadrangles were unprojected to geographic coordinates, appended to one another, and edge-matched. Selected shorelines, lakes, and rivers were added from the hydrographic coverage. The geology was reconciled along the borders of the adjacent quadrangles, and the entire map was then converted to the Lambert azimuthal equal area projection.
8. Errors in polygon labeling were checked by using standard ArcInfo routines.

Following peer review, the database, checkplot, and map unit descriptions were reviewed by a map editor and a map layout was prepared. Production procedures for this map generally followed those described by Lane and others (1999): importing ArcInfo shapefiles for the geologic database and planimetric base into Adobe Illustrator through the MAPublisher plugin (Avenza software); assigning line styles, colors, and patterns through the "Select by Attribute" function of MAPublisher; and adding text, figures, and marginalia in Illustrator. For this map, polygons were labeled by using the "Feature Text Label" function of MAPublisher.

Our greatest obstacle in the production of the map layout was the "spider web" effect that resulted from importing the polygons in the ESRI shapefiles into Adobe Illustrator through MAPublisher. This distortion of the imported polygons is the result of some polygon boundaries having too many vertices for Illustrator to handle. We first split the original polygon coverage into two coverages, north and south, but this splitting did not eliminate the "spider web" effect. We considered generalizing the boundaries in ArcInfo, but it might have changed the linework in unpredictable and unacceptable ways. Instead, we chose to construct grids and superimpose them on the polygons of the problem coverages. Cutting the polygons up in this way simplified their boundaries. Arcs in the new coverages had fewer vertices, and we could import them into Adobe Illustrator without generating "spider web" distortions.

The procedure involves the following two steps, which may be repeated as needed to break up the polygons (because of the complexity of the polygon coverages, we found it necessary to impose a series of grids on the original polygons):

1. Create a grid coverage, drawing the lines of the grid so they intersect arcs that have the most vertices.
2. UNION the grid coverage with the original polygon coverage to create a modified polygon coverage.

The modified polygon coverages actually consisted of more polygons than did the original coverages. However, each part of an original polygon that was intersected by a gridline was attributed the same as the parent polygon. After importing the modified polygon coverages into Illustrator, we used the "Select by Attribute" function of the MAPublisher plugin to assign colors and patterns to the polygons, and to label the units.

Dividing the original coverage into two areas, north and south, required us to select most of the 185 map units twice in order to assign colors, patterns, and unit labels—twice the work of these operations on one coverage.

SOFTWARE CITED

Adobe Illustrator—Adobe Systems Inc., 345 Park Ave., San Jose, CA 95110-2704, (408)536-6000, <http://www.adobe.com>.

ArcInfo—Environmental Systems Research Institute (ESRI) Inc., 380 New York St., Redlands, CA 92373-8100, (909) 793-2853, <http://www.esri.com>.

LT4X—Infotec Development GIS Products & Services, 500 NE Multnomah, Suite 329, Portland, OR, 97232.

MAPublisher—Avenza Systems Inc., 6505-B Mississauga, Ontario, CANADA L5N 1A6, <http://www.avenza.com>.

REFERENCES

Fullerton, D.S., Bush, C.A., and Pennell, J.N., 2003, Map of surficial deposits and materials in the Eastern and Central United States (east of 102° west longitude): U.S. Geological Survey Geologic Investigations Series Map I–2789, 1 sheet, scale 1:2,5000,000; pamphlet, 48 p., <http://pubs.usgs.gov/imap/I-2789/>.

Lane, D.E., Donatich, A., Brunstein, F.C., and Shock, N., 1999, Digital geologic map production and database development in the Central Publications Group of the Geologic Division, U.S. Geological Survey, *in* Soller, D.R., ed., Digital Mapping Techniques '99—Workshop Proceedings (May 1999, Madison, Wisc.): U.S. Geological Survey Open-File Report 99–386, p. 11–15., <http://pubs.usgs.gov/of/of99-386/lane.html>.

APPENDIX A

List of Workshop Attendees

[Grouped by affiliation]

Alaska Geological Survey
Carrie L. Browne

Arizona Geological Survey
Stephen M. Richard

Bryn Mawr Geology Department
Maria Luisa Crawford

Colorado State University/National Park Service
Stephanie O'Meara
Trista Thornberry

Delaware Geological Survey
Lillian Wang

EMS Environment Institute
Brian Bills

Geologic Mapping Institute
Michael W. Higgins
Ralph Crawford

Geological Survey of Alabama
Gene Daniel Irvin

Geological Survey of Canada
Eric Boisvert
Martin Anctil

GeoReference Online Ltd.
Clinton Smyth

Idaho Geological Survey
Jane Freed
Loudon R. Stanford

Illinois State Geological Survey
Brett Denny
Daniel Byers
Jane Domier
John McLeod
Melony Barrett
Rob Krumm
Sheena Beaverson

Iowa Geological Survey
Huaibao P. Liu

Kansas Geological Survey
Jorgina A. Ross
William E. Harrison

Kentucky Geological Survey
Jerry Weisenfluh

Louisiana State University
Robert Paulsell

Millersville University—Earth Sciences
Alex DeCaria
Lynn Marquez

National Park Service
Anne R. Poole

Natural Resources Canada
Andrew Moore
Boyan Brodaric
Dave Everett
Mike Sigouin
Victor Dohar

New Jersey Geological Survey
Ronald S. Pristas

New Mexico Bureau of Geology
David J. McCraw
Glen Jones

North Carolina Geological Survey
Jeffrey C. Reid
John G. Nickerson
Mark W. Carter

North Dakota Geological Survey
Lorraine A. Manz
Mark A. Gonzalez

Ohio Geological Survey
Donovan M. Powers
James McDonald
Thomas M. Berg

Oregon Department of Geology
Paul E. Staub

Pennsylvania Geological Survey
Christina Miles
Gale Blackmer
Gary Fleeger
Jay Parrish
Michael Moore
Rodger Faill
Stuart Reese
Thomas G. Whitfield
William E. Kochanov

South Carolina Geological Survey
Erin Hudson
Scott Howard

U.S. Forest Service
Andrew H. Rorick

U.S. Geological Survey
Alex Donatich
Bruce R. Johnson
Charles A. Bush
Christopher P. Garrity
David R. Soller
Diane Lane
Jonathan C. Matti
Joseph East
Nancy Stamm
Pat Leahy
Peter Schweitzer
Peter T. Lyttle
Robert S. Wardwell

University of Alabama
Douglas Behm

University of California, Santa Barbara
Jordan T. Hastings

University of New Brunswick
Chris Parsons
Wouter Van de Poll

Utah Geological Survey
Kent Brown

Virginia Division of Mineral Resources
Ian Duncan

Washington Division of Geology and Earth Resources
Charles Caruthers

West Virginia Geological Survey
Gayle H. McColloch, Jr.
Jane S. McColloch

Wisconsin Geological Survey
Kurt Zeiler

☆ U.S. GOVERNMENT PRINTING OFFICE: 2004 — 675-241/ 05002 Region No. 8