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

Digital Mapping Techniques '06—Workshop Proceedings

Edited by David R. Soller

June 11-14, 2006
Columbus, Ohio

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

Hosted by the
Ohio Geological Survey

U.S. GEOLOGICAL SURVEY OPEN-FILE REPORT 2007-1285
2007

CONTENTS

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 '06 (DMT'06) workshop was attended by more than 110 technical experts from 51 agencies, universities, and private companies, including representatives from 27 state geological surveys (see Appendix A of these Proceedings). This workshop was similar in nature to the previous nine meetings, which were held in Lawrence, Kansas (Soller, 1997), Champaign, Illinois (Soller, 1998), Madison, Wisconsin (Soller, 1999), Lexington, Kentucky (Soller, 2000), Tuscaloosa, Alabama (Soller, 2001), Salt Lake City, Utah (Soller, 2002), Millersville, Pennsylvania (Soller, 2003), Portland, Oregon (Soller, 2004), and Baton Rouge, Louisiana (Soller, 2005). This year's meeting was hosted by the Ohio Geological Survey, from June 11-14, 2006, on the Ohio State University campus in Columbus, Ohio. As in the previous meetings, the objective was to foster informal discussion and exchange of technical information. It is with great pleasure that I note that the objective was successfully met, as attendees continued to share and exchange knowledge and information, and renew friendships and collegial work begun at past DMT workshops.

Each DMT workshop has been coordinated by the Association of American State Geologists (AASG) and U.S. Geological Survey (USGS) Data Capture Working Group, the latter of 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, this volume, and http://ngmdb.usgs.gov/info/standards/datacapt/). The Working Group was formed because increased production efficiencies, standardization, and quality of digital map products were needed for the database—and for the State and Federal geological surveys—to provide more high-quality digital maps to the public.

At the 2006 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) field data capture software and techniques, including the use of LIDAR; 3) digital cartographic techniques; 4) migration of digital maps into ArcGIS Geodatabase format; 5) analytical GIS techniques; and 6) continued development of the National Geologic Map Database.

ACKNOWLEDGEMENTS

I thank the Ohio Geological Survey (OGS) and their Chief and State Geologist, Thomas M. Berg, for hosting this meeting and for arranging for corporate sponsorship. During the past 10 years, I have worked closely with the Association of American State Geologists and, in particular, with Tom, who retired soon after the DMT'06 meeting. I wish to thank him profusely for his many years of guidance, support, and friendship.

In the tradition of past DMT meetings, the attendees were given a very informative, productive, and enjoyable experience. I especially thank Jim McDonald (OGS), who coordinated the events. Other OGS staff who deserve thanks are those who provided essential support for this meeting, including Lisa Van Doren (for preparing graphics, signs, and maps), Mac Swinford (poster boards), Ed Kuehnle (poster boards), Madge Fitak (registration), Sharon Stone (meeting logistics), Garry Yates, and Dennis Hull (poster boards).

The meeting was co-hosted by The Ohio State University Department of Geological Sciences, and I thank them for their hospitality and for their significant contributions to this meeting. Specifically, I thank Franklin Schwartz (Department Chair and meeting sponsor), Karen Tyler (facilities setup), Garry McKenzie (general assistance), Ken Shelberg (finances, registration), Sue Shipley (finances, registration), Mary Scott (Sunday reception host), Dale Gnidovec (Sunday reception host), Michael Seufer (web site), Betty Heath (registration), Mary Hill (registration), and Kelley Barrett, Kelley Carroll, Mike Fidler, Steve Goldsmith (setup, registration), and Brent Curtis (WiFi setup).

The meeting was greatly improved through the generous financial donations of the National Office and the Ohio Chapter of the American Institute of Professional Geologists (AIPG), the Ohio Industrial Minerals and Aggregates Association, and the Ohio Oil and Gas Association. The Eastern Section of AAPG and the West Virginia Geological Survey generously provided the poster boards, and for this we owe them a warm thank you.

I also thank the members of the Data Capture Working Group (Warren Anderson, Kentucky Geological Survey;

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; George Saucedo, California Geological Survey; and Tom Whitfield, Pennsylvania Geological Survey) for advice in planning the workshop's content.

I warmly thank Lisa Van Doren (Ohio Geological Survey) for typesetting the Proceedings. And, last but not least, I thank all attendees for their participation; their enthusiasm and expertise were the primary reasons for the meeting's success.

PRESENTATIONS AND POSTERS

The workshop included 32 oral presentations and 25 posters. Many 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 having to "reinvent the wheel."

During the course of the 10 annual DMT meetings, it has been my pleasure to meet, and work with, the many talented people who have authored papers in these Proceedings. As the subjects addressed by the DMT meetings have become even more essential to the Nation's geological surveys, the demands placed on them have risen to the point where many authors scarcely have time to address their work fully. Predictably, less time is then available to compose written summaries of their work; I'm sure the readers (or at least other editors) can sympathize with this predicament. Therefore, I include with this Introduction a list of all presentations and posters (Appendix A of this paper). If the reader finds an interesting title that isn't recorded in these Proceedings, I encourage them to contact the authors directly. Further, some presentations and related information is available for download at http://ngmdb.usgs.gov/Info/dmt/DMT06presentations.html.

THE NEXT DMT WORKSHOP

The eleventh annual DMT meeting will be held in the Spring of 2007, on the campus of South Carolina University, in Columbia, South Carolina. Please consult the Web site (http://ngmdb.usgs.gov/Info/dmt/) for updated information. While planning for that event, the Data Capture Working Group will carefully consider recommendations for meeting content and format offered by DMT'06 attendees.

REFERENCES

Soller, D.R., editor, 2005, Digital Mapping Techniques '05—Workshop Proceedings: U.S. Geological Survey Open-File Report 2005-1428, 268 p., accessed at http://pubs.usgs.gov/of/2005/1428/.

Soller, D.R., editor, 2004, Digital Mapping Techniques '04—Workshop Proceedings: U.S. Geological Survey Open-File Report 2004-1451, 220 p., accessed at http://pubs.usgs.gov/of/2004/1451/.

Soller, D.R., editor, 2003, Digital Mapping Techniques '03—Workshop Proceedings: U.S. Geological Survey Open-File Report 03-471, 262 p., accessed at http://pubs.usgs.gov/of/2003/of03-471/.

Soller, D.R., editor, 2002, Digital Mapping Techniques '02—Workshop Proceedings: U.S. Geological Survey Open-File Report 02-370, 214 p., accessed at 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., accessed at 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., accessed at http://pubs.usgs.gov/of/of00-325/.

Soller, D.R., editor, 1999, Digital Mapping Techniques '99—Workshop Proceedings: U.S. Geological Survey Open-File Report 99-386, 216 p., accessed at http://pubs.usgs.gov/of/of99-386/front.html.

Soller, D.R., editor, 1998, Digital Mapping Techniques '98—Workshop Proceedings: U.S. Geological Survey Open-File Report 98-487, 134 p., accessed at http://pubs.usgs.gov/of/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., accessed at http://pubs.usgs.gov/of/of97-269/.

APPENDIX A

List of oral and poster presentations, and discussion sessions.

Oral Presentations

A 10-year retrospective on the Digital Mapping Techniques workshops
By David R. Soller (U.S. Geological Survey)

The new Geology Ontario web portal - an "out-of-the-box" solution for discovering and delivering Ontario's geoscience data
By Zoran Madon (Ontario Geological Survey)

Building a water well database for GIS analysis
By A. Wayne Jones and Kelly A. Barrett (Ohio Department of Natural Resources, Division of Water)

Identifying sensitive aquifers in Ohio
By Chris Kenah, Michael Slattery, Linda Slattery, and Michael Eggert (Ohio EPA)

Discussion Session - "Topographic maps and framework data in the future"
This session focused on plans for creating and updating topographic and other framework map data, through local, state, and national partnerships. The session began with these presentations:

- Enhancing USGS topo quads, and GIS for the Gulf, by Stafford Binder (U.S. Geological Survey)
- Building NSDI through local, state, and national partnerships, by Stu Davis (National States Geographic Information Council (NSGIC)
- The National Map, by Charles Hickman (U.S. Geological Survey)

GeoSciML, a GML application for geoscience information interchange
By the CGI Data Model and Testbed working group

Open source web-mapping, the Oregon experience
By David Percy (Portland State University)

High Resolution DEM's from digital photogrammetry, stereo-autocorrelation, and morphological filtering as an alternative to LIDAR for mapping applications
By Peter G. Chirico (U.S. Geological Survey)

Integration of high-resolution satellite imagery for coastal mapping and monitoring
By Ron Li, Xutong Niu, Sagar Deshpande, Feng Zhou, and Kaichang Di (The Ohio State University)

GIS in use at an industrial minerals company
By Steve Murdoch (Oglebay Norton Company / O-N Minerals)

Discussion Session on LIDAR
This session focused on LIDAR technology, image processing techniques, and its application to geologic mapping. The session began with these presentations:

- LIDAR basics,by Jim Giglierano (Iowa Geological Survey)
- Airborne Laser Swath Mapping (LIDAR) and geology: The B4 project, by Michael Bevis, David Raleigh, Shan Shan, Dana Caccamise, Eric Kendrick, and Wendy Shindle (The Ohio State University), Ken Hudnut (U.S. Geological Survey), and Dorota Grejner-Brzezinska and Charles Toth (The Ohio State University)
- LIDAR and various levels of accuracy. by Mark Brooks (Optimal Geomatics)

The National Park Service Geologic Resources Evaluation; Subtitle: "Using GIS to get GIS"
By Timothy B. Connors (National Park Service)

The concept and development of the National Geological Map Database in the Czech Republic
By Robert Tomas (Czech Geological Survey)

Geological Map Database - A practitioner's guide to delivering the information
By Jeremy Giles (British Geological Survey)

The Publishing Process Integration system for the Geological Survey of Canada publication products
By Linda Guay (Geological Survey of Canada)

Building Geodatabase coded-value domains from National Geologic Map Database vocabularies
By Steve Richard (Arizona Geological Survey), and Jon Craigue and Dave Soller (U.S. Geological Survey)

Visualizing earthquake hazard information in Google Earth
By J. Luke Blair, Marco Ticci, James Lienkaemper, and Heather Lackey (U.S. Geological Survey)

Preserving North Carolina legacy geologic and topographic maps
By Jeffrey C. Reid (North Carolina Geological Survey), Jeff Essic (North Carolina State University Libraries), Steve Morris (North Carolina State University Libraries), and Smitha Ramakrishnan (University of North Carolina, Greensboro)

Saturation and value modulation: A new method for integrating colour and grey-scale imagery
By David Viljoen and Jeff Harris (Geological Survey of Canada)

Geographic Imager software for Adobe Photoshop
By David Andrec and Doug Smith (Avenza Systems, Inc.)

Acquiring unpublished geologic evidence to augment Web dissemination of Kentucky's geologic maps
By Jerry Weisenfluh (Kentucky Geological Survey)

The challenges and benefits of distributing digital data: Lessons learned
By Kenneth Papp, Susan Seitz, and Larry Freeman (Alaska Division of Geological & Geophysical Surveys), and Carrie Browne (formerly with ADG&GS)

The Alabama Metadata Portal: A new solution for serving large amounts of data
By Philip Patterson (Geological Survey of Alabama)

IT Security - How it affects digital mapping
By Harry McGregor (University of Arizona and the U.S. Geological Survey)

From Geek to Illiterate Manager: following the road wherever it leads, enjoying the scenery and ignoring the address ranges
By Jay Parrish (Director, Pennsylvania Bureau of Topographic and Geologic Survey)

3D geological modeling: Solving a classification problem with the Support Vector Machine
By Alex Smirnoff, Eric Boisvert, and Serge J. Paradis (Geological Survey of Canada)

Qualitative and quantitative 3D modeling of surficial materials at multiple scales
By Erik Venteris (Ohio Geological Survey)

Discussion Session - "A vision for web-accessible 3D geological mapping"
Led by Harvey Thorleifson (Director, Minnesota Geological Survey), this session offered for discussion this vision and how it might be implemented:

"People require geological mapping to fulfill their objectives related to health, heritage, safety, and economic development, and they expect public information to be web-accessible and readily usable. Could the geoscience community cooperate in order to make available an international database that provides known and predicted subsurface conditions, based on consistent global coverage, zoomable down to the most detailed coverage available, with links to the source map? To address societal issues, the database should include bathymetry, soils, onshore and offshore surficial and bedrock geology, and 3D geology depicting extent, thickness, and properties of geologic units, so that web-accessible drillhole forecasts can be issued for any point."

Poster Presentations

Geologic map of the Ouachita Mountain region in Arkansas
By Boyd R. Haley and Charles G. Stone (Arkansas Geological Commission)

Template for a geologic map at 1:24,000 scale
By William D. Hanson and Jerry W. Clark (Arkansas Geological Commission)

A Geodatabase schema for geologic map production
By Vic Dohar (Geological Survey of Canada)

The Publishing Process Integration system for the Geological Survey of Canada publication products
By Linda Guay (Geological Survey of Canada)

The art of mapping with a catalogue of geo-knowledge: Sable Island Bank and the Gully, Scotian Shelf, offshore eastern Canada
By Edward L. King and Gary M. Grant (Geological Survey of Canada)

Saturation and value modulation: A new method for integrating colour and grey-scale imagery
By David Viljoen and Jeff Harris (Geological Survey of Canada)

Digital map production at the Czech Geological Survey, Czech Republic
By Zuzana Krejci (Czech Geological Survey)

Compression of digital orthophotography collections
By Deette Lund (Illinois State Geological Survey)

Recent LGS StateMap geologic maps; Recent LGS geologic lithographs; The impact of Hurricane Katrina
By R. Hampton Peele, Richard P. McCulloh, Paul Heinrich, John Snead, Lisa Pond, Robert Paulsell, DeWitt Braud, Ahmet Binselam, Ivor van Heerden, and Rob Cunningham (Louisiana Geological Survey and Louisiana State University)

Surficial and 3-D geological mapping in support of land and water management in Manitoba, Canada
By Greg Keller and Gaywood Matile (Manitoba Geological Survey)

Converting Adobe Illustrator maps to ArcMap format
By Jennifer Mauldin (Nevada Bureau of Mines and Geology)

Preserving North Carolina legacy geologic and topographic maps
By Jeffrey C. Reid (North Carolina Geological Survey), Jeff Essic (North Carolina State University Libraries), Steve Morris (North Carolina State University Libraries), and Smitha Ramakrishnan (University of North Carolina, Greensboro)

Using GIS to create and analyze potentiometric-surface maps
By Paul N. Spahr, A. Wayne Jones, Kelly A. Barrett, Michael P. Angle, and James M. Raab (Ohio Department of Natural Resources, Division of Water)

Detailed, three-dimensional, surficial-geology mapping of the Milan, Ohio 1:24,000 Quadrangle
By Rick Pavey (Ohio Geological Survey)

Updates to the Known and Probable Karst Map of Ohio
By Donovan Powers (Ohio Geological Survey)

New map of the surficial geology of the Lorain and Put-In-Bay 30 x 60 Minute Quadrangles, Ohio
By E.M. Swinford, R.R. Pavey, G.E. Larsen, and K.E. Vorbau (Ohio Geological Survey)

Airborne Laser Swath Mapping (LIDAR) and Geology: The B4 project
By Michael Bevis, David Raleigh, Shan Shan, Dana Caccamise, Eric Kendrick, and Wendy Shindle (The Ohio State University), Ken Hudnut (U.S. Geological Survey), and Dorota Grejner-Brzezinska and Charles Toth (The Ohio State University)

The National Park Service Geologic Resources Evaluation; Subtitle: “Using GIS to get GIS”
By Timothy B. Connors (National Park Service)

GIS and GPS utility in the geologic mapping of complex geologic terrane on the Mascot, Tennessee 7.5’ Quadrangle
By Barry W. Miller and Robert C. Price (Tennessee Division of Geology)

Spatial adjustment and digital capture of unprojected geologic data for the USGS 2004 oil and gas assessment of the Michigan Basin
By Joseph A. East (U.S. Geological Survey)

Prototype GIS database for the DNAG Geologic Map of North America
By Christopher Garrity and David Soller (U.S. Geological Survey)

Publications Warehouse: A database of verified, Web-enabled citations, USGS publications, and their metadata (http://pubs.usgs.gov)
By Carolyn McCullough and Greg Allord (U.S. Geological Survey)

The National Geologic Map Database
By David R. Soller (U.S. Geological Survey), Thomas M. Berg (Ohio State Geologist), and Nancy R. Stamm (U.S. Geological Survey)

USGS National Surveys and Analysis projects: Preliminary compilation of integrated geological datasets for the United States
By Doug Stoeser, Ric Wilson, Steve Ludington, Connie Dicken, and Suzanne Nicholson (U.S. Geological Survey)

Banding birds with MapServer
By Rob Wardwell and Kevin Laurent (U.S. Geological Survey)

The National Geologic Map Database Project: Overview and Progress

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

In the past decade, the National Geologic Map Database (NGMDB) project has evolved from a general concept to a set of resources that have helped the Nation's geological surveys provide to the public, in a more efficient manner, standardized digital geologic map information. Throughout this period, I have had the honor of serving as the NGMDB project chief. In this capacity, I worked closely with the Association of American State Geologists (AASG; see Appendix A) and, in particular, with my AASG chief liaison, Thomas M. Berg (State Geologist and Chief, Ohio Geological Survey). Tom retired soon after the DMT'06 meeting, and I wish to thank him profusely for his many years of guidance, support, and friendship.

After 10 years of managing the NGMDB project and organizing the Digital Mapping Techniques workshops, I found that a slight reduction in certain activities has been necessary in order to bring new priorities and direction to the project. For example, by comparison to the progress reports of previous years (see Appendix B), this report is quite limited in scope. Below, I briefly document the project's progress during this year; for detailed descriptions of this project's goals and accomplishments, please refer to the DMT'05 report (Soller and others, 2005).

BACKGROUND

Development and management of science databases for support of societal decisionmaking and scientific research are critical and widely recognized needs. The National Geologic Mapping Act of 1992 and its subsequent reauthorizations stipulate creation and maintenance of a National Geologic Map Database (NGMDB), as a national archive of spatially referenced geoscience data, including geology, paleontology, and geochronology. The Act further stipulates that all new information contributed to the NGMDB adhere to standards, which are to be developed as needed under the guidance of the NGMDB project. Development of a national database and its attendant standards is a daunting task that requires close collaboration among all geoscience agencies in the U.S., at the State and Federal levels. The Act, therefore, creates the environment within which the USGS and the Association of American State Geologists can collaborate to build the NGMDB and serve as well the specific needs of their own agencies.

From the guidelines in the National Geologic Mapping Act, and through extensive discussions and forums with the geoscience community and with the public, a general strategy was designed by which to build the National Geologic Map Database (NGMDB). This strategy was publicly stated in 1995 and, based on public input, has gradually evolved. The NGMDB is designed to be a comprehensive reference tool and data management system for spatial geoscience information in paper and digital form. More specifically, the NGMDB will consist of the following: 1) limited metadata in its Map Catalog for all paper geoscience maps and book publications that contain maps (including maps of any part of the Nation, published by any agency), online viewable images of paper and digital maps, metadata for published digital map data, and links to online data; 2) ancillary databases that provide further information about geoscience in the USGS and the state geological surveys (e.g., the Geologic Names Lexicon, the Mapping in Progress Database, and the National Paleontology Database); 3) nationwide geologic map coverage at intermediate and small scales; 4) an online database of geologic maps (predominantly in vector format; planned as a distributed system); 5) a set of Web interfaces to permit access to these products; and 6) a set of standards and guidelines to promote more efficient use and management of spatial geoscience information. The NGMDB system is a hybrid – some aspects are centralized and some are distributed, with the map information held by various cooperators (for example, the State geological surveys). Through a primary entry point on the Web, users can browse and query the NGMDB, and obtain access to the information wherever it resides.

Project Organization

The project has been designed as a set of related tasks that will develop, over time, a NGMDB with increasing complexity and utility. This is being accomplished through a network of geoscientists, computer scientists, librarians, and others committed to supporting the objectives of the NGMDB. ***Phase One*** of this project principally involves the building of a comprehensive Map Catalog of bibliographic records and online images of all available paper and digital maps, and book publications containing maps, that adhere to the earth-science themes specified in the National Geologic Mapping Act of 1992. The first phase also includes the design and development of the Geologic Names Lexicon, the Mapping in Progress Database, and the National Paleontology Database. ***Phase Two*** is the development of standards and guidelines for geologic map and database content and format. ***Phase Three*** is a long-term effort to develop a database that contains nationwide geologic map coverage at a variety of map scales, according to a complex set of content and format specifications that are standardized through general agreement among all partners in the NGMDB (principally the AASG); this database will be integrated with the databases developed in Phase One. The NGMDB project's technology and standards development efforts also are coordinated with various international bodies, including the Federal Geographic Data Committee, the North American Geologic Map Data Model Steering Committee, ESRI, the U.S. National Science Foundation's Geoinformatics project "GEON," the IUGS Commission on the Management and Application of Geoscience Information ("IUGS CGI"), the IUGS Commission on Stratigraphy, the IUGS-affiliated Commission for the Geological Map of the World, and the International Association of Mathematical Geology (IAMG).

A full realization of the project's third phase is not assured and will require a strong commitment among the cooperators as well as adequate technology, map data, and funding. The project will continue to assess various options for development of this database, based on realistic funding projections and other factors. During the development of these phases of the NGMDB, extensive work will be conducted to generate Web interfaces and search engines, and continually improve them. Development of the data management and administrative protocols will be a priority as well, to ensure that the NGMDB will function efficiently in the future. The NGMDB's databases and project information can be found at http://ngmdb.usgs.gov.

PROGRESS

Phase One

As noted above, the objective of Phase One is to provide quick access to existing geoscience resources. As evidence of success in this activity, the NGMDB "Phase One" databases (Map Catalog, Geolex, Mapping in Progress) receive about 135,000 visits a month from about 35,000 users (nearly all of whom are non-USGS). This usage has increased dramatically since mid-2005 – most likely this was achieved through: 1) increased content; 2) persistent (and more effective?) outreach to users at scientific meetings and through email and publications; 3) a successful appeal to USGS to identify NGMDB as the link from "Geologic Information" on the USGS home page; and 4) increased market prices for earth resource commodities, which thereby increases the demand for geoscience information. With this increased Web traffic has come an increase in user requests for information and assistance – these users vary widely in interest and background, and include school children, homeowners, local government planners, and professional geologists. With many of these users we have personal contact by email to ensure they find what they need.

Specific accomplishments this year include:

1. Expansion of the NGMDB Map Catalog by about 4000 records, to a total of about 75,000 records. This includes 36,000 USGS maps in map, book, and open-file series, essentially all relevant USGS publications. It also includes 26,000 State survey publications and 13,000 products by other publishers.
2. Engagement of 45 states in the process of entering Map Catalog records and processing of about 2000 new records for State geological survey publications.
3. Doubling of the number of links to online publications, including map images; the total is now 10,000.
4. Continued the expansion and revision of Geolex (the geologic names lexicon), with a major update completed in mid-year.
5. Completion of the incorporation of the prototype Image Library into the Map Catalog. The Image Library utilized a subset of Catalog records and provided a Web interface that did not easily scale upward to accommodate new images. This incorporation was a significant effort that will provide

users with a more productive search process; the project is now focused on providing a geographic-search capability within the Map Catalog to give users a search option we attempted to provide via the Image Library.

6. Writing of a prototype application that generates a file to display Map Catalog search results in Google Earth. This application was made available for public comment.
7. Completion of several hundred productive interchanges with Map Catalog, Geolex, and Image Library users via the NGMDB feedback form and other mechanisms.
8. Numerous project presentations to scientists and managers at USGS, AASG, and other scientific meetings, whereby details of the project were explained and participation of professionals in building various NGMDB standards and databases was increased (e.g., Map Catalog, Geolex, online map database).
9. Participation with USGS National Cooperative Geologic Mapping Program (NCGMP) in an effort to begin to revise significantly the Mapping in Progress database, focusing on database redesign and adding information that will be useful to NCGMP management.
10. The providing of index maps, in response to requests by USGS and AASG management, that show areas in the U.S. that have been geologically mapped at various scales and time periods (see Soller, 2005). These maps and statistics (e.g., square miles mapped at 1:24,000-scale from year 2000 to 2005) were presented at various venues and were used by NCGMP to prepare responses to the Office of Management and Budget during their annual performance appraisal.

Phase Two

Phase Two addresses a Congressional mandate to develop standards and guidelines for geologic map and database content and format. Specific accomplishments this year by members of the NGMDB project staff include:

1. Coordination of work on the Federal Geographic Data Committee's draft standard for geologic map symbolization; revisions to the standard, based on FGDC Standards Working Group review; gaining final approval from the FGDC Coordination Committee for release as the Federal standard. This standard includes: a new terminology for representing the scientific and locational confidence associated with geologic map features (e.g., contacts, faults, sample locations), a Postscript implementation of the standard (the ArcGIS implementation is under development), and a comprehensive response to all comments received in the FGDC-sanctioned Public Review.
2. Serving as Chair of the FGDC Geologic Data Subcommittee.
3. Organization and leadership of the tenth annual "Digital Mapping Techniques" workshop, and publication of the workshop Proceedings from the previous year's meeting (DMT '05, Baton Rouge, LA). These meetings have proven to be a principal means by which to help the geoscience community converge on more standardized approaches for digital mapping and GIS analysis.
4. Serving as committee Secretary and as member of the newly-formed U.S. Geologic Names Committee, and assistance in proposal of geologic time scale and color scheme for adoption by USGS.
5. Serving as Coordinator of the North American Geologic Map Data Model Steering Committee (NADMSC) and managing the NADM website.
6. Continuing to provide strong intellectual input on design and implementation of the NADMSC conceptual data model ("NADM C1.0"). This data model was published in late 2004 and is based in part on results of the NGMDB-Kentucky database prototype that was developed in 2002-03.
7. Serving as U.S. representative to DIMAS, the global standards body that serves the Commission for the Geological Map of the World. Provision of technical information and guidance on data model and science language standards under development in North America, and participation in DIMAS initiatives to develop global standards.
8. Serving as the U.S. Council Member to the IUGS Commission for the Management and Application of Geoscience Information ("CGI").
9. Participation in the IUGS CGI-sponsored "International Data Model Collaboration Working Group." Assistance in developing consensus for international standards for a geologic data model. Contributing to development of the XML-format "GeoSciML" schema, which will be proposed as an international data-exchange standard.
10. Serving as IUGS CGI liaison to the "Multi-Lingual Thesaurus Working Group." This group is

enabling global exchange of geoscience information by developing a common science vocabulary that is translated into many languages.

11. Continued interaction with ESRI, regarding collaboration on an ArcGIS Geology Data Model.
12. Serving as member of IAMG Strategic Planning Committee and providing guidance regarding GIS and IT issues.

Phase Three

As noted above, Phase Three is a long-term effort to develop a database that contains nationwide geologic map coverage at a variety of map scales, according to a complex set of content and format specifications that are standardized through general agreement among all partners in the NGMDB. Project activities this year included:

1. Continued development of the prototype database, focusing on compilation of a standard science terminology, implementation of the NADM conceptual data model in ESRI's ArcGIS, and creation of a data-entry tool to assist geologists and GIS specialists in creating standardized map databases. The prototype data model was posted to the ESRI Geology Data Model website for evaluation by the international community.
2. Participation in the Database Interoperability Testbed #2, sponsored by the IUGS CGI's Working Group on Data Model Collaboration. This testbed required disparate map data to be managed in a prototype online map database system that could demonstrate various query and symbolization functionality as well as the ability to output selected map data to the GeoSciML data interchange format. A critical part of this task was identifying and contracting for highly skilled geologists with strong backgrounds in programming, GIS, spatial database design, and Web delivery of information. This is a vitally important testbed involving at least 8 agencies worldwide. NGMDB participation involved the Arizona Geological Survey, Portland State University, DOGAMI (Oregon GS), and the University of Arizona.
3. In order to have modern, small-scale, consistent geologic map coverage for the U.S., the NGDMB project is converting the recently published Geologic Map of North America (GMNA) to digital format. This is a daunting task, and so an area was selected in which a prototype map database would be developed (it included part of the U.S., Canada, and the Pacific Ocean). The prototype map was created and subjected to peer review at the DMT'06 meeting (see Garrity and Soller, this volume). This prototype demonstrated the feasibility of converting the enormously complex map files from Adobe Illustrator to ArcGIS. Participating agencies (Geological Society of America, Geological Survey of Canada, Woods Hole Oceanographic Institute, and USGS) have since been contacted with regard to finalizing the NGMDB proposal so as to create and manage the GMNA map database.

ACKNOWLEDGEMENTS

I principally thank my long-time colleague on this project, Thomas M. Berg (State Geologist and Chief, Ohio Geological Survey), for his friendship and innumerable contributions to the success of this project. I also thank the NGMDB project staff and collaborators for their enthusiastic and expert support, without whom the project would not be possible. In particular, I thank: Nancy Stamm (USGS, Reston; Geolex database manager and associate project chief); Ed Pfeifer, Alex Acosta, Dennis McMacken, Jana Ruhlman, and Michael Gishey (USGS, Flagstaff and Tucson, AZ; Website and database management), Chuck Mayfield (USGS; Map Catalog content), Robert Wardwell and Ben Carter (USGS, Vancouver, WA, and Reston, VA; Image Library), Steve Richard (Arizona Geological Survey, Tucson, AZ; data model and science terminology), Jon Craigue (University of Arizona /USGS, Tucson, AZ; data-entry tool), and David Percy (Portland State University; Google Earth application for the Map Catalog). I also thank the many committee members who provided technical guidance and standards (Appendix A).

REFERENCES

Soller, D.R., 2005, Assessing the Status of Geologic Map Coverage of the United States—A New Application of the National Geologic Map Database, in D.R. Soller, ed., Digital Mapping Techniques '05—Workshop Proceedings: U.S. Geological Survey Open-file Report 2005-1428, p.41-47, available at http://pubs.usgs.gov/of/2005/1428/soller2/.

Soller, D.R., Berg, T.M., and Stamm, N.R., 2005, The National Geologic Map Database Project: Overview and Progress, in D.R. Soller, ed., Digital Mapping Techniques '05—Workshop Proceedings: U.S. Geological Survey Open-file Report 2005-1428, p. 23-40, available at http://pubs.usgs.gov/of/2005/1428/soller1/.

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)
- Larry Becker (Vermont Geological Survey)
- Rick Berquist (Virginia Division of Mineral Resources)
- Jim Cobb (Kentucky Geological Survey)
- Ian Duncan (Texas Bureau of Economic Geology)
- 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)

Geologic Data Subcommittee of the Federal Geographic Data Committee:
- Dave Soller (U.S. Geological Survey and Subcommittee Chair)
- Jerry Bernard (USDA-Natural Resources Conservation Service)
- Mark Crowell (Dept. of Homeland Security, Federal Emergency Mgmt. Agency)
- Jim Gauthier-Warinner (U.S. Forest Service, Minerals and Geology Management)
- Laurel T. Gorman (U.S. Army Engineer Research and Development Center)
- John L. LaBrecque (National Aeronautics and Space Administration)
- Lindsay McClelland (National Park Service)
- Jay Parrish (State Geologist, Pennsylvania Geological Survey)
- George F. Sharman (NOAA National Geophysical Data Center)
- Dave Zinzer (Minerals Management Service)

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 Working Group Chair)
- Warren Anderson (Kentucky Geological Survey)
- Rick Berquist (Virginia Geological Survey)
- 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

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)
- Peter Davenport (Geological Survey of Canada)
- Bruce Johnson (U.S. Geological Survey and Chair of the Data Interchange Technical Team)
- Rob Krumm (Illinois State Geological Survey)
- Scott McColloch (West Virginia Geological and Economic Survey)
- Steve Richard (Arizona Geological Survey)
- Loudon Stanford (Idaho Geological Survey)
- Jerry Weisenfluh (Kentucky Geological Survey)

IUGS Commission for the Management and Application of Geoscience Information:
- Dave Soller (U.S. Geological Survey, Council Member)

Conceptual Model/Interchange Task Group (of the Data Model Collaboration Working Group of the IUGS Commission for the Management and Application of Geoscience Information):
- Steve Richard (Arizona Geological Survey, Task Group Member)

DIMAS (Digital Map Standards Working Group of the Commission for the Geological Map of the World):
- Dave Soller (U.S. Geological Survey, Working Group Member)

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://ngmdb.usgs.gov/info/statecontacts.html for this list.

These groups have fulfilled their mission and are no longer active:

NGMDB 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)

AASG/USGS Metadata Working Group:

Peter Schweitzer (U.S. Geological Survey and Working 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 Working 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 Working 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)

APPENDIX B

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

Soller, D.R., editor, 2005, Digital Mapping Techniques '05—Workshop Proceedings: U.S. Geological Survey Open-File Report 2005-1428, 268 p., accessed at http://pubs.usgs.gov/of/2005/1428/.

Soller, D.R., editor, 2004, Digital Mapping Techniques '04—Workshop Proceedings: U.S. Geological Survey Open-File Report 2004-1451, 220 p., accessed at http://pubs.usgs.gov/of/2004/1451/.

Soller, D.R., editor, 2003, Digital Mapping Techniques '03—Workshop Proceedings: U.S. Geological Survey Open-File Report 03-471, 262 p., accessed at http://pubs.usgs.gov/of/2003/of03-471/.

Soller, D.R., editor, 2002, Digital Mapping Techniques '02—Workshop Proceedings: U.S. Geological Survey Open-File Report 02-370, 214 p., accessed at 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., accessed at 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., accessed at http://pubs.usgs.gov/of/of00-325/.

Soller, D.R., editor, 1999, Digital Mapping Techniques '99—Workshop proceedings: U.S. Geological Survey Open-File Report 99-386, 216 p., accessed at http://pubs.usgs.gov/of/of99-386/front.html.

Soller, D.R., editor, 1998, Digital Mapping Techniques '98—Workshop Proceedings: U.S. Geological Survey Open-File Report 98-487, 134 p., accessed at http://pubs.usgs.gov/of/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., accessed at http://pubs.usgs.gov/of/of97-269/.

Soller, D.R., Berg, T.M., and Stamm, N.R., 2005, The National Geologic Map Database Project: Overview and Progress, *in* D.R. Soller, ed., Digital Mapping Techniques '05 – Workshop Proceedings: U.S. Geological Survey Open-File Report 2005-1428, p. 23-40, accessed at http://pubs.usgs.gov/of/2005/1428/soller1/.

Soller, D.R., Berg, T.M., and Stamm, N.R., 2004, The National Geologic Map Database project: Overview and progress, *in* Soller, D.R., ed., Digital Mapping Techniques '04—Workshop Proceedings: U.S. Geological Survey Open-File Report 2005-1451, p. 15-31, accessed at http://pubs.usgs.gov/of/2004/1451/soller/.

Soller, D.R., and Berg, T.M., 2003, The National Geologic Map Database project: Overview and progress, *in* Soller, D.R., ed., Digital Mapping Techniques '03—Workshop Proceedings: U.S. Geological Survey Open-File Report 03-471, p. 57-77, accessed at http://pubs.usgs.gov/of/2003/of03-471/soller1/.

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, accessed at 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, accessed at 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, accessed at http://pubs.usgs.gov/of/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, accessed at http://ngmdb.usgs.gov/info/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, accessed at http://pubs.usgs.gov/of/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, accessed at http://pubs.usgs.gov/of/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, accessed at http://ngmdb.usgs.gov/info/reports/geotimes97.html.

Soller, D.R., and Berg, T.M., 1995, Developing the National Geologic Map Database: Geotimes, v. 40, no. 6, p. 16-18, accessed at http://ngmdb.usgs.gov/info/reports/geotimes95.html.

The New *Geology Ontario* Web Portal—An "out-of-the-box" Solution for Discovering and Delivering Ontario's Geoscience Data

By Zoran Madon[1], P.Geo. and Brian Berdusco[2], M.Sc.

[1]Ontario Geological Survey (OGS)/Ministry of Northern Development and Mines
Willet Green Miller Centre, B7
933 Ramsey Lake Road
Sudbury, ON P3E 6B5
Telephone: (705) 670-5991
Fax: (705) 670-5905
e-mail: zoran.madon@ontario.ca

[2]Business Solutions/Land and Resources Cluster
Ministry of Northern Development and Mines (MNDM)
159 Cedar Street, Suite 605
Sudbury, ON P3E 6A5
Telephone: (705) 564-7093
Fax: (705) 564-7919
e-mail: Brian.Berdusco@ontario.ca

INTRODUCTION

The new *Geology Ontario* web portal (www.ontario.ca/geology) was developed during the last two years as a direct result of the Minister's Office directive to improve the discovery and delivery of Ministry of Northern Development and Mines (MNDM) geoscience data (Figure 1). This directive came about in response to clients' complaints about MNDM's existing delivery mechanisms. Many clients work in mining jurisdictions throughout North America and around the world, and are accessing geoscience data in a variety of ways. Not surprisingly, they quickly adopt the best methods for accessing data and, subsequently, demand this high level of service delivery from other jurisdictions.

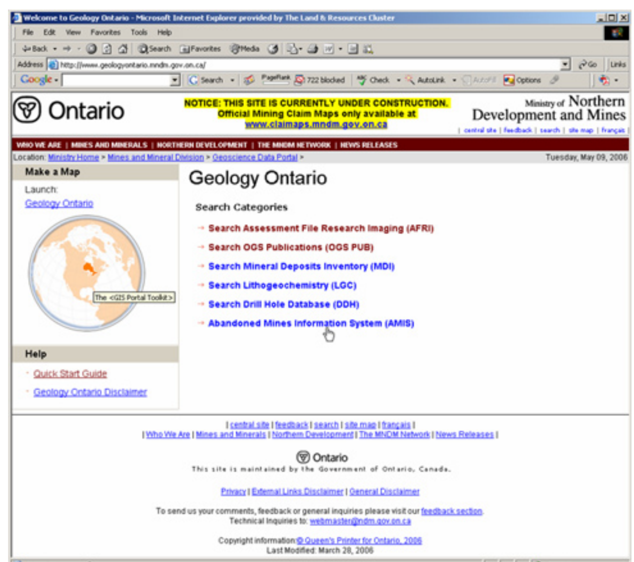


Figure 1. ***Geology Ontario*** home page allowing access to both the map search window as well as the text search windows for the MNDM's geoscience data archives (AFRI, PUBS, MDI, LGC, ODHDB and AMIS).

BACKGROUND

The GeoPortal Project was initiated in the spring of 2004. The project team, made up of a project manager and a small group of individuals seconded from various sections within MNDM, held several meetings initially to scope out the requirements and to assign responsibilities.

The project specifications were constrained by the following core requirements:

- create a delivery website using an "out-of-the-box" solution that meets current needs
- concentrate on delivering all OGS publications (PUBS), mineral exploration assessment files (AFRI), the Mineral Deposit Inventory (MDI), the Ontario Drill Hole Database (ODHDB), the Lithogeochemical Database (LGC), land tenure information (*CLAIMaps III*), and Abandoned Mines Information (AMIS) – totaling in excess of 300 gigabytes of data including 2 million text pages and 168,000 maps.

Work on the *Geology Ontario* web portal began on several fronts, including the selection of a software vendor to provide the web service, a benchmarking study, a client survey, development of appropriate metadata records for all data holdings, and conversion of various existing digital image files to a more useful, popular and web-friendly format. This initial work was followed up with business-impact and threat-risk studies, focus group sessions, and a Quality Assurance/Quality Control process to test the web portal functionality.

Software Vendor Selection Process

In 1997, MNDM consolidated nine regional Provincial Mining Recorder Offices into a single, centralized office in Sudbury, Ontario. This necessitated a new way to distribute 40,000 mining land tenure (claim) maps annually, as well as a new process for maintaining information on these maps. The Ministry decided to use the internet as a service delivery mechanism, and as a result, the Crown Land Automated Internet Mapping System (*CLAIMaps*) was developed. *CLAIMaps*, originally a simple ArcView IMS application, delivered scanned images of claim maps over the internet. However, the process of updating these maps was limited to editing hard copies, rescanning them, and then reposting them on the Ministry's website.

In 2000, the decision was made to update the CLAIMaps application based on Environmental Systems Research Institute (ESRI) Map Objects Internet Map Service (MOIMS) software. The new application, CLAIMaps II, quickly grew in use, and the heightened demand placed on the system resulted in critical, technical challenges on server infrastructure. In addition, the solution was based on complex custom code that required significant human and financial resources to maintain. This system ran from March 2001 to December 2002.

The next phase of the application was built on an out-of-the-box philosophy, and a new *CLAIMaps* website based on ESRI ArcSDE, ArcIMS, and Oracle RDBMS was launched. The new system, *CLAIMaps III*, turned out to be a stable application that required few resources to maintain and resulted in a greatly reduced total cost of ownership. It was also recognized internationally as a world class system capable of generating and delivering daily-updated map images and polygon data (Figure 2).

The success of CLAIMaps III made ArcSDE, ArcIMS, and Oracle the obvious choices as the base for building *Geology Ontario*. The Ministry approached companies on the provincial Vendor of Record and requested a time and materials quote for developing the new portal. Companies had to show a proficiency in developing web-based applications based on ESRI ArcSDE, ArcIMS, Oracle RDBMS, and the ArcIMS Portal Toolkit. Like *CLAIMaps*, the development environment for *Geology Ontario* was the Java 2 Plaform, Enterprise Edition (J2EE).

Benchmarking Study

Benchmarking is a crucial step in helping define customer requirements, establishing objectives, developing productivity measures, and remaining competitive. In this benchmark study, over a hundred geoscience web sites were examined in order to document industry best practices and ensure that the *Geology Ontario* web portal would provide the necessary services to meet client requirements. Data discovery methods (i.e., map search tools, text search tools), data content, and data format were tabulated for each site visited.

Over 50% of the sites offer text-based search tools, while more than 30% offer map-based search tools. Approximately 25% of these use ArcIMS technology. The rest maintain either simple listings in some order (i.e., chronological, alphabetical, by subject matter) because of the low volume of downloadable data or do not offer any data for download, but simply provide a link to their publications office. Most of the sites with downloadable data have some form of metadata, and many of these are FGDC/Z39.50 compliant. Some sites even produce thumbnail images of the document or map.

Almost all of the downloadable sites supply geoscience publications, both maps and reports, as well as regional thematic data that encompasses their entire jurisdiction. These would typically include regional planimetric, geological and geophysical maps, mineral deposit and regional geochemical data, and in some cases regional terrain models (DEMs) and satellite imagery. Some sites offer mineral land tenure maps and assessment file reports, as well.

By far, the most popular format for disseminating both reports and maps is the Adobe Systems PDF (Portable Document Format). GIS and vector data downloads are available in various formats, but the ESRI Arcview/ArcInfo formats predominate. Tabular data stored in spreadsheets is generally provided in Microsoft Excel and ASCII CSV formats. Very few sites supply image and grid data

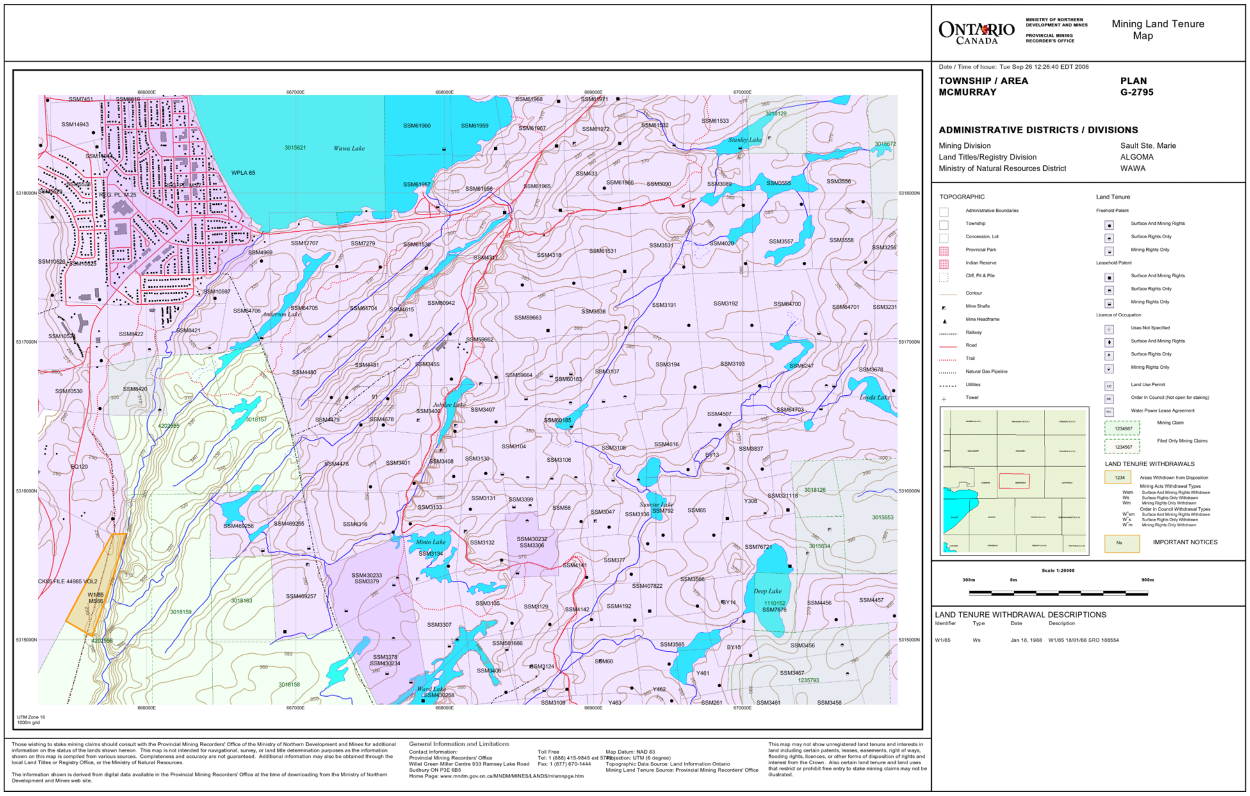


Figure 2. Sample of a typical map product generated from the *CLAIMaps III* web site.

(satellite images, geophysical data), so no dominant format emerged here. Rather, they appear to be spread evenly across the following types – Lizardtech MrSID, Geosoft GRD/GXF, ERMapper ERS, TIFF, JPG, and BMP.

All sites, with the exception of one, provide free downloads. Some sites require individuals to register online before being allowed to download, particularly sites that provide access to very large files such as geophysical data and satellite imagery. Disclaimers and liability waivers vary from the non-existent to those that appear before each download. Some sites offer the possibility of downloading data in more than one format and, for maps, in more than one projection. Most sites provide only current data for download – generally less than 5 to 10 years old.

Client Survey

A client survey was posted on MNDM’s e-consultation web site for approximately 6 weeks in the summer of 2004. The survey consisted of 26 questions, primarily in multiple choice format, and generally took less than 10 minutes to complete. Throughout the questionnaire, respondents were given the opportunity to comment, critique, and provide suggestions. The survey was subdivided into the following sections:

- background information (business activity, geographic area of interest, internet connection used, etc.),
- data discovery/access preferences (map-based vs. text-based searches, text-based search criteria),
- data content/format preferences (which government geoscience data sets were most important, what formats do you use for report data, map data, GIS data, tabular data, etc.)
- data usage (plotting on maps, incorporating into reports, spatial modeling, reference material, etc.) and
- data quality issues (method for reporting data errors, etc.).

Approximately 200 respondents completed the survey, which provided the project team with a better profile of clients’ needs and preferences as well as a wealth of constructive suggestions to incorporate into the web design. Suggestions included the following requests: list examples or provide pick lists for all text-search criteria, develop an on-line help utility, produce hyperlinks to documents directly from the search results page, and provide an easy link to communicate problems or issues pertaining to the web portal.

Metadata Development and Data Conversion Process

Metadata pertaining to MNDM publications is currently housed in two different databases. In 1994, the Ministry developed a database that contains attributes on all publications produced since 1891. In 2002, the Ministry began entering metadata into a centralized metadata server based on the Government of Ontario Information Technology Standard (GOITS) 72.0. GOITS 72.0 is a metadata standard that was derived from the U.S. Federal Geographic Data Committee (FGDC) standards. Though new efforts have concentrated on the centralized metadata server, MNDM continues to populate the older publications database simply because this database contains records for all publications.

The Ministry has created a process of extracting metadata from the publications database and populating the metadata catalog available in ArcIMS. In this implementation, MNDM has utilized the FGDC compliant catalog but will review moving to the new ISO standard in the near future. The metadata server is a Z39.50 protocol compliant server that enables direct access from library search engines.

Metadata values were also used to populate fields within the PDF documents themselves. The purpose of populating these fields or meta-tags is to facilitate the rapid discovery of pertinent information either through the website search engine or through other common search engines such as Google and Yahoo. These search engines not only index content, but also index information contained within PDF meta-tags. The Ministry also embedded meta-tag data within the mineral assessment files to facilitate discovery.

The majority of publications and assessment files were scanned in the early 1990s and stored as raster images. MNDM utilized raster to PDF conversion software available from JRAPublish and converted approximately 2,000,000 pages and 168,000 maps to 85,000 PDF documents. These documents, in turn, were subjected to optical character recognition software and converted to “PDF searchable” format. The Ministry built a 40,000 entry geoscientific and geographic names dictionary to enhance word pattern recognition by the conversion software. The net result is that the majority of these millions of raster images can now be searched for content using the dtSearch engine that has been implemented on the *Geology Ontario* website.

Simple and complex Boolean functions coupled with meta-tag search fields ensure that clients are provided with simple tools that return powerful results (Figure 3). For example, using the following Boolean syntax: *visible w/3 gold* returns those occurrences where the word *visible* is found within three words of the word *gold*. Though this may seem like an obvious query to perform, what must be realized is that performing this type of query will also return occurrences of *no visible gold*. A simple modification of the above query to: *visible w/3 gold not w/2 no* will remove the occurrences of *no visible gold*. Combining these Boolean queries with meta-tag data can assist the client in narrowing down results and pinpointing searches.

In the future, metadata will probably be entered, stored, and managed through ESRI ArcCatalog. Not only will searches align with the Z39.50 protocol, but the data distributed through the website will also include necessary attributes for proper spatial queries. ArcCatalog also conforms to the out-of-the-box philosophy that has contributed to a reduction in the total cost of the development and maintenance of internet-facing, GIS-related web applications.

Business-Impact Analysis and Threat-Risk Analysis

As work on *Geology Ontario* progressed, the project team undertook a business-impact analysis, a threat-risk analysis, and a business continuity plan, using guidelines developed by the Ontario government’s business audit group.

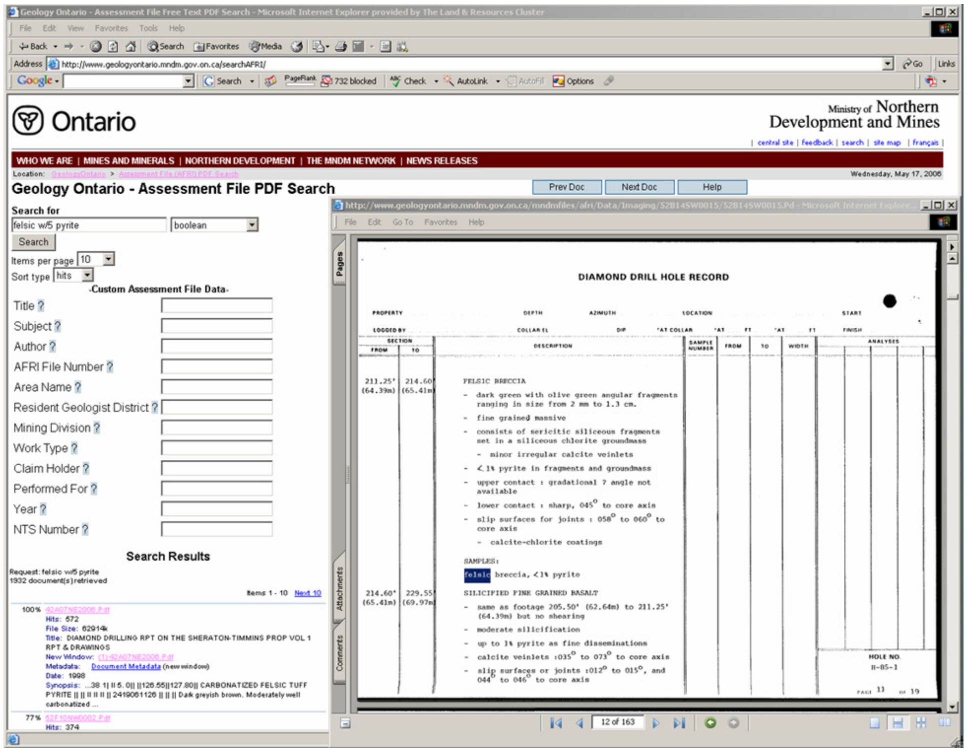
The business impact study attempts to evaluate how an interruption in the *Geology Ontario* web service would impact clients over time. Although several client impact variables are available in the guidelines (including public and employee health and safety, business service delivery, public confidence, internal perception, regulatory and legal, environmental impact, financial reporting, revenue loss), it was concluded that disruption of the web service would have minimal impact for only business service delivery in the short term (< 1 day) and moderate to high impact for public confidence and internal perception in the longer term (2 weeks to 1 month).

The threat-risk analysis evaluated all the *Geology Ontario* assets, including data, software/hardware architecture, network, and staff. For each asset, a sensitivity assessment with respect to confidentiality, integrity, availability, and accountability was estimated to determine the potential severity of harm if the asset was compromised in any way. Threat agents, both internal (i.e., users, system administrators) and external (i.e., clients, hackers, contractors, natural and man-made disasters, etc.), as well as threat events, were identified. Once determined, an estimate for the likelihood (i.e., low, moderate or high) of each threat event as well as its exposure (i.e., the impact to the government if the threat is realized) was established.

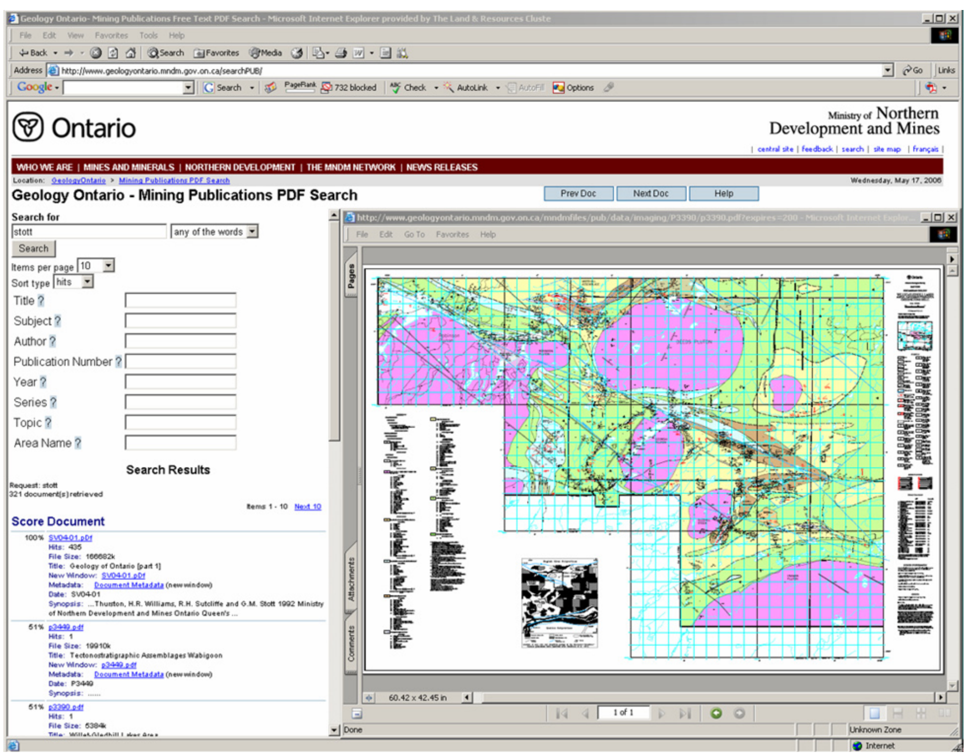
Finally, a business continuity plan was developed to recommend alternate strategies for delivering *Geology Ontario* services during an unexpected interruption.

Focus Group

When a beta version of the website was available for review and testing, three individuals who represented



A



B

Figure 3. ***Geology Ontario*** "PDF searchable" files allow for advanced search capability; A) in this example, a Boolean search for the word "felsic" within 5 words of "pyrite" returned almost 2000 hits from the ***Geology Ontario*** AFRI archives; B) in this example, a simple search using an OGS geologist's last name returned over 300 documents, including maps from the ***Geology Ontario*** PUBS archive.

MNDM’s main client groups were invited to Sudbury for a “test drive.” After a few sessions, the individuals completed a Ministry evaluation form to provide feedback on all aspects of the *Geology Ontario* web portal, including:

- the portal home page functionality, layout, and content
- on-line help
- layout and functionality of the search pages for the various themes
- layout and functionality of the map viewer window
- overall search functionality
- layout and functionality of the search results window
- functionality of the download process
- general comments when navigating between different windows.

Their input helped to further refine and improve the ***Geology Ontario*** web portal.

Quality Assurance and Quality Control

The *Geology Ontario* website replaces the Earth Resources Mineral Exploration webSite (ERMES). However, since both *Geology Ontario* and ERMES operate on copies of the same Oracle databases, quality assurance processes were developed whereby query results generated in ERMES were compared with the same results generated in *Geology Ontario*. Any discrepancies between the two pointed to possible inherent problems within query structures. In some cases, it was discovered that the original queries in ERMES were incorrect.

Data integrity issues were a persistent problem in both the OGS Publications and AFRI databases. Historically, files were loaded using a complex data loading process developed in the early 1990s, which involved populating Oracle databases and storing raster images of individual pages and maps. It was not uncommon for data loading problems to cause orphaned attribute or image data. Orphaned documents could not be queried and would remain, essentially, undiscoverable. However, once these scanned images were converted to searchable PDF format and indexed, they became discoverable through the text search engine. Orphaned files have now been identified and targeted for correction so that both attribute and text-based queries return similar results.

The original scanned documents from the early 1990s were stored in either JPG or TIFF format, and as thumbnail GIF images. On occasion, the original TIFF images were scanned and stored incorrectly, which rendered them unconvertible to PDF. MNDM worked with the PDF software vendor (James Rile Associates) and developed a process whereby a specially designed page was inserted in place of any unusable TIFF image. Embedded within the “bad page” notice was a special code that was, in turn, indexed by the text search engine. By searching for this code, a record was generated that indicated which file pages were troublesome. The “bad page” notice, therefore, serves the dual purpose of informing clients that a page was scanned incorrectly and failed to convert, and assisting the development and maintenance team in identifying which pages required re-scanning.

JPG images that were created at that time failed to populate the data pertaining to resolution (dots per inch or dpi) in the JPG header. The PDF conversion application defaulted to an incorrect dpi which, when the JPGs were converted, caused them to display at the wrong scale. Since the original scanning was done at 200 dpi for JPG images, an image conversion utility was run to update the headers in every JPG file with the 200 dpi setting. As a result, these images converted to the correct scale in the final PDF document.

The Ministry will develop a problem report page where clients can submit information on any problems they encounter with products downloaded from the ***Geology Ontario*** website. In addition, clients can recommend changes that would assist them in using the web-based products.

GEOLOGY ONTARIO WEB PORTAL

Benefits of Out-of-the-Box Software

The entire application development environment adhered to the same “out-of-the-box” philosophy that has made the *CLAIMaps* website so successful. Minimal custom code generation will facilitate the Ministry’s ability to manage and modify the *Geology Ontario* website on an ongoing basis with the limited resources that the Ministry has at its disposal. Other benefits to an “out-of-the-box” solution include:

- large pool of skills available to select from (i.e. software vendor)
- limited programming expertise required
- with version upgrades, solution systems can easily implement new functionality
- puts pressure on software vendors to enhance “out-of-the-box” functionality (as opposed to custom coding)
- develops common skill sets and solution methodologies across an organization.

In January 2003, the *CLAIMaps III* website was launched by the Mineral Development and Lands Branch. This application utilizes ESRI ArcSDE and ArcIMS technologies and ORACLE 8i RDBMS together with Macromedia JRun 3.1 to create maps that are served through Microsoft Internet Information Server on a Microsoft Windows 2000 Server. The *CLAIMaps* application website, which provides 3 million map images and 70,000 final maps per year, has garnered international recognition and awards by delivering daily updated land tenure maps-to-scale for Ontario.

The graphical user interface of the *CLAIMaps* website is intuitive and easy to use, which are key selling points with MNDM's client base. In addition, clients can link into the ArcIMS map services and incorporate land tenure data into their private, proprietary database systems to generate mining land tenure maps and other complex analytical products in the privacy of their work or home environments.

The *CLAIMaps* website is updated every night through the Automated CLAIMap Management System (ACMS) Replication manager software synchronizes land tenure databases between the internal production server (ACMS) and the public facing internet server (CLAIMaps) thus fulfilling the additional function of providing offsite storage of the Ministry's data. If required, these data can readily be used for business resumption in the event of catastrophic failure of either the ACMS or *CLAIMaps* servers.

The Ministry's key clients have clearly indicated that they would prefer the delivery of the other Ministry products, specifically geoscience data and reports, through an application that operates in a fashion comparable to the *CLAIMaps* website.

Out-of-the-Box Software

The new ***Geology Ontario*** portal website was built using the latest versions of the same software utilized by the *CLAIMaps* website, namely, Oracle 9i, ArcSDE 9, ArcIMS 9, XML, java, JavaScript, and Apache Tomcat (J2EE). Data is served through the Internet Information Server component available in Microsoft Server 2003. The Oracle and ArcSDE software will reside on the database server, while ArcIMS and Windows Internet Information Server will reside on a separate application server.

The *CLAIMaps* website, which currently resides on a Windows 2000 server, will be moved to the *Geology Ontario* servers. All environment variables will be modified so that the *CLAIMaps* application points to the new Oracle/ArcSDE server. The ArcSDE component of the Oracle database will be tuned to optimize performance. The data that reside within the current Oracle 8i databases will be migrated to Oracle 9i. The replication process that currently updates the *CLAIMaps* server will be revised to point to the new Oracle/ArcSDE server so that data can be updated in a similar fashion. The critical components of the new *Geology Ontario* website include:

- Land Information Ontario
- Ontario Land Information Directory
- Internet Geospatial Data Delivery
- XML, ArcXML, Java, Javascript
- ESRI ArcIMS 9, AcSDE 9, ArcGIS 9
- ESRI ArcMap Server, ArcGIS Server, Portal Toolkit
- ESRI Metadata Server, Java ADF
- Apache Tomcat
- Oracle 8i migration to 9i
- Ontario Mining Land Tenure Data
- Geoscience Digital Data
- *CLAIMaps III* Functionality
- ArcIMS
- Windows 2003 Server
- iSERV Production Facilities and the processes.

Highlights

The guiding principle driving this project was to enhance discovery and download of all MNDM geoscience data, maps, reports, and publications. In excess of 300 gigabytes of data have been made available on the *Geology Ontario* web portal. All existing assessment files and publications (over 2 million pages and 168,000 maps), stored as image files, were converted to "PDF searchable" format for more rapid and convenient downloading. As mentioned, the "PDF searchable" format allows for powerful data mining capabilities, since all PDF documents are indexed and can be queried for virtually any text string or combination of text strings. Also, the "PDF searchable" format makes it possible to incorporate government data into a client's work environment using copy/paste functionality (Figure 4). Once downloaded onto a client's personal computer, these documents can be indexed and their contents discoverable using desktop search tools like Google Desktop and Copernic.

Data discovery is based on the ESRI Portal Toolkit and metadata tools (Figure 5). Clients enter search criteria, generate a search, and receive the results in a tabulated format with a description of the documents and their file sizes as well as a hotlink for direct download (Figure 6). In addition, *Geology Ontario* provides links to other Ministry web sites (Figure 7) and map services, which allows clients simultaneous access to services of other government agencies, including those in other Provinces (Figure 8).

For digital data products, such as CAD drawings, ArcView shapefiles, and database files, clients will be required to accept a disclaimer prior to download. The disclaimer is in the form of an html/asp document, controlled by the Internet Information Server component of Windows 2003 server. The Government of Ontario is limited with respect to placing cookies on client computers and, as a result, the disclaimer acceptance avoids leaving cookies.

The Ministry has also developed a map browsing site, much like the *CLAIMaps* website, where clients can simultaneously view numerous thematic datasets (Figure 9). Clients have clearly indicated their preference for viewing

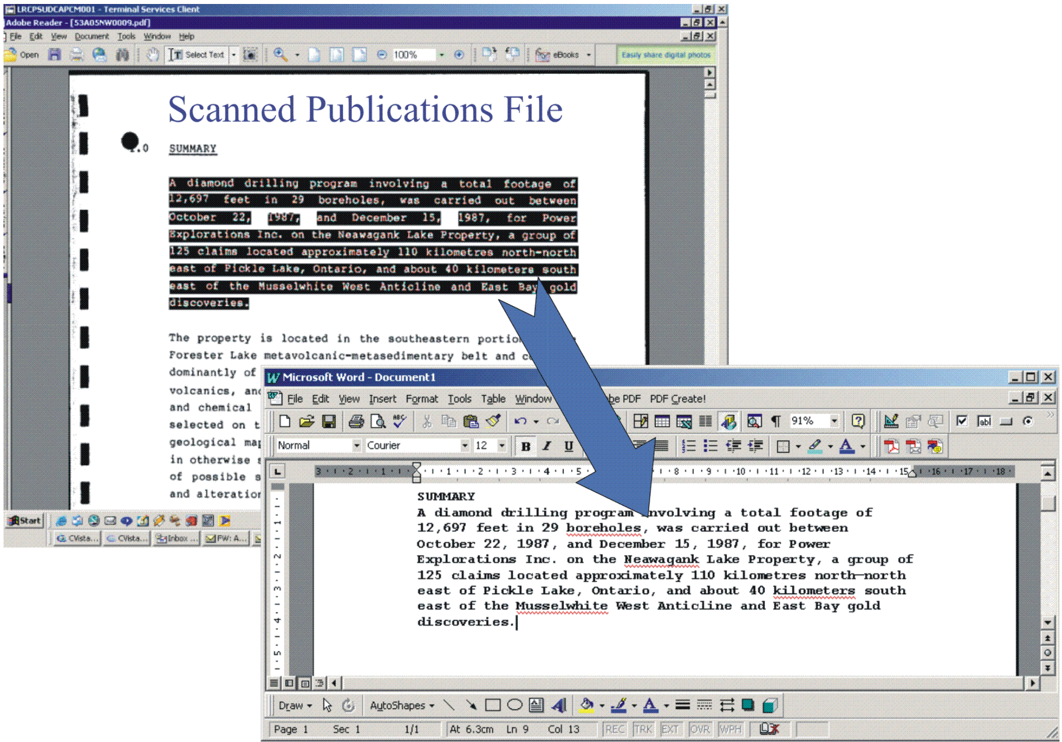


Figure 4. ***Geology Ontario*** "PDF searchable" files allow for copy and paste functionality with all its downloadable documents.

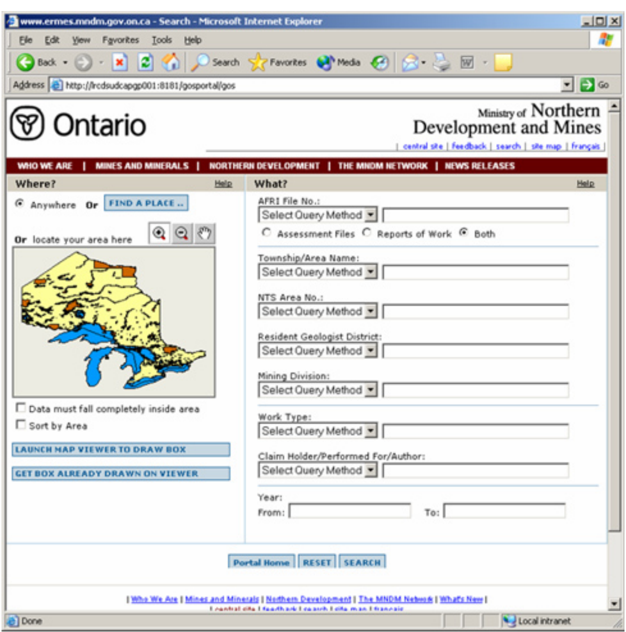


Figure 5. ***Geology Ontario*** text search window for querying the mineral exploration assessment files database (AFRI).

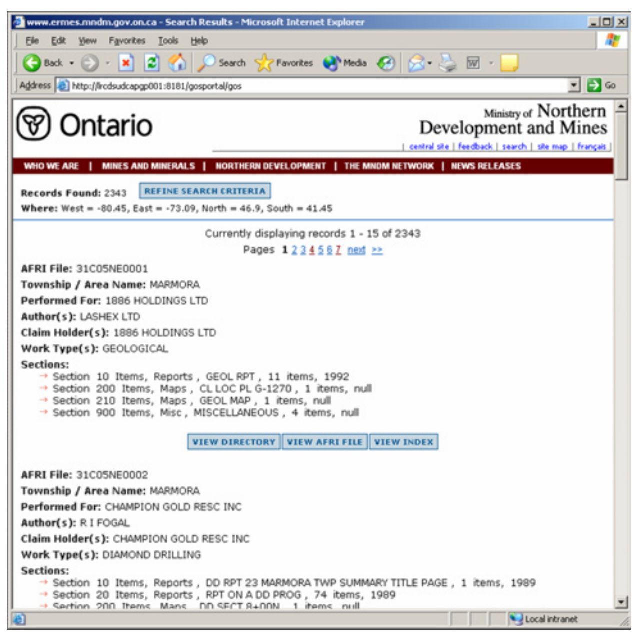


Figure 6. Search results from a typical query of the AFRI database with linkages to detailed metadata and to the folder containing the downloadable files.

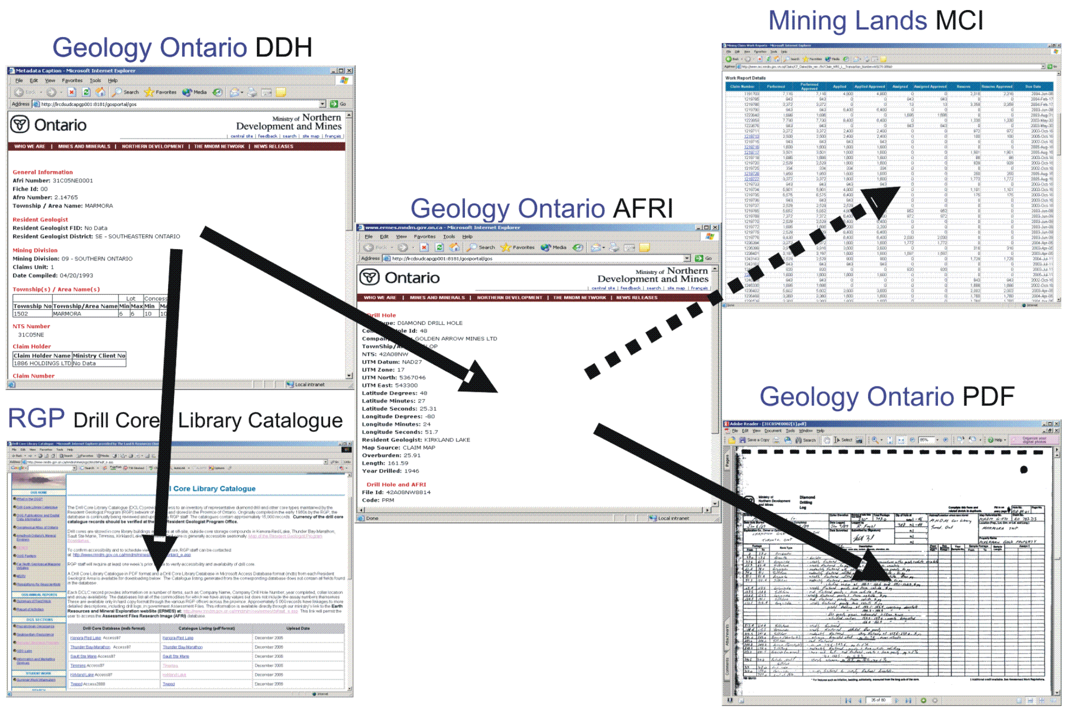


Figure 7. The ***Geology Ontario*** web site provides links to other Ministry web sites – either directly using hypertext (solid arrows) or indirectly using reference numbers in the metadata (dashed arrow). (abbreviations: DDH – diamond drill hole database; RGP – Resident Geologist Program; AFRI – Assessment File Research Imaging database; MCI – Mining Claims Information database; PDF – Portable Document Format).

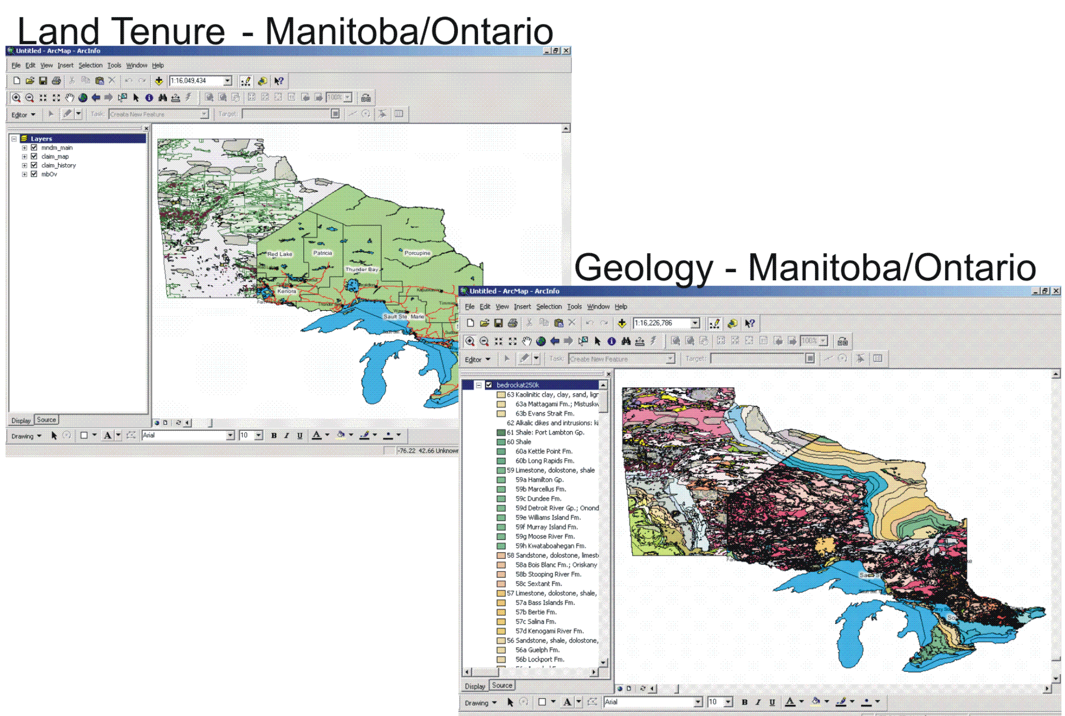


Figure 8. Because ***Geology Ontario*** uses ArcIMS technology, clients can simultaneously link to ArcIMS-based web sites from other jurisdictions – in this illustration land tenure and geology are being accessed concurrently from both the ***Geology Ontario*** web portal and the Manitoba portal.

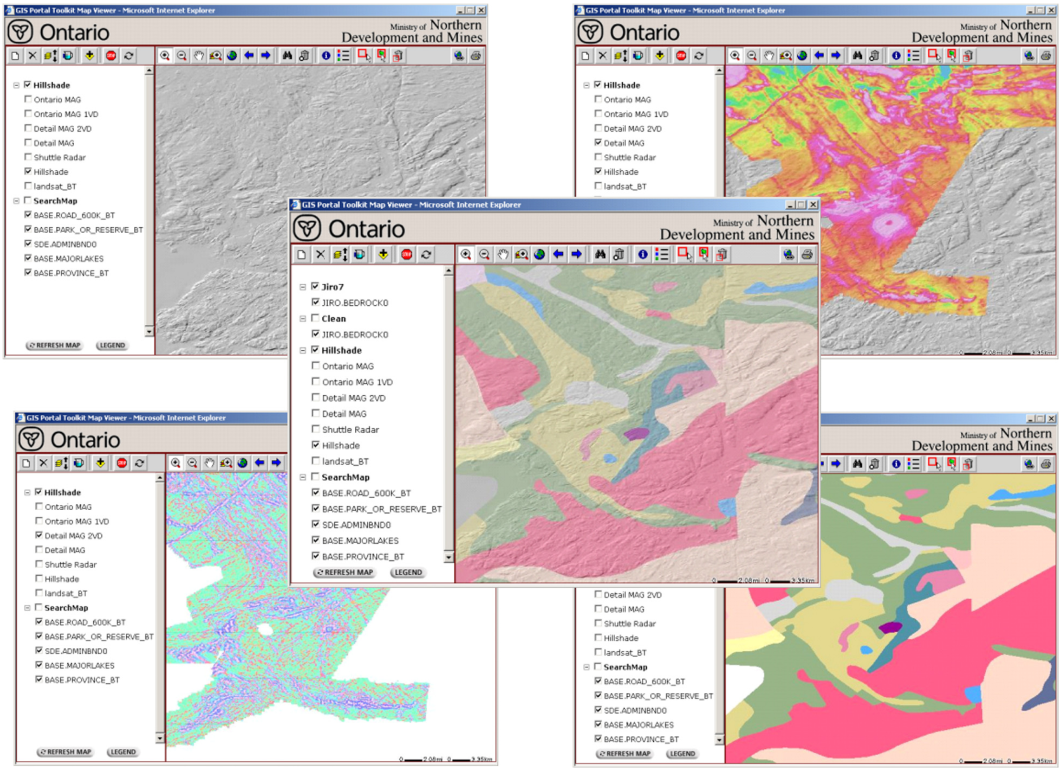


Figure 9. Samples of the ***Geology Ontario*** map browser window illustrating clockwise from top left, a DEM, total field magnetics draped over a DEM, geology, magnetics, and geology draped over a DEM (center image). Over 2 dozen pan-Ontario themes are available for viewing in the map browser window.

mineral land tenure together with geology, geophysics, geochemistry, and other thematic data in order to relate land tenure to known geoscientific characteristics. The GeoPortal Project Team has created geological, geophysical, geochemical, and geomorphologic thematic datasets for multi-theme viewing. Clients can also print maps to scale in a fashion comparable to the *CLAIMaps* website and include various geoscience themes in the map output.

FUTURE WEB DELIVERY

In addition to delivering geoscience data to MNDM clients worldwide, the Ministry is also looking at enhancing the visual web delivery of its data. Some avenues being explored include using 3-dimensional perspective tools with multi-thematic overlay capabilities through a standard web browser (Figure 10) and reviewing best practices in use by other organizations, as far as delivering geoscience data through interfaces such as Google Earth (Figure 11), Microsoft Virtual Earth, NASA World Wind, or ESRI ArcReader and ArcGIS Explorer. Utilizing these additional applications requires careful planning and a balanced client-server approach. Large, static datasets such as geophysical grids digital elevation and remotely sensed imagery can be downloaded and reside on the client side while dynamic datasets such as drilling data and rock sample analyses can be accessed directly from the *Geology Ontario* server through keyhole markup language (kml) or Open Geospatial Consortium (OGC) web mapping services.

ACKNOWLEDGEMENTS

The authors would like to thank Terry Johnston, a member of the *Geology Ontario* project team, for her critical review of this paper.

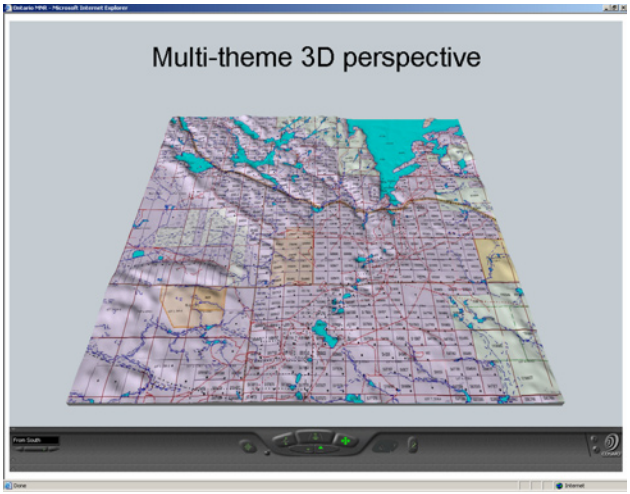


Figure 10. Possible future web viewing enhancements for ***Geology Ontario*** – providing 3-dimensional perspective tools with multi-thematic overlay capabilities through a simple web browser; in this illustration, the mineral land tenure fabric is draped over a DEM.

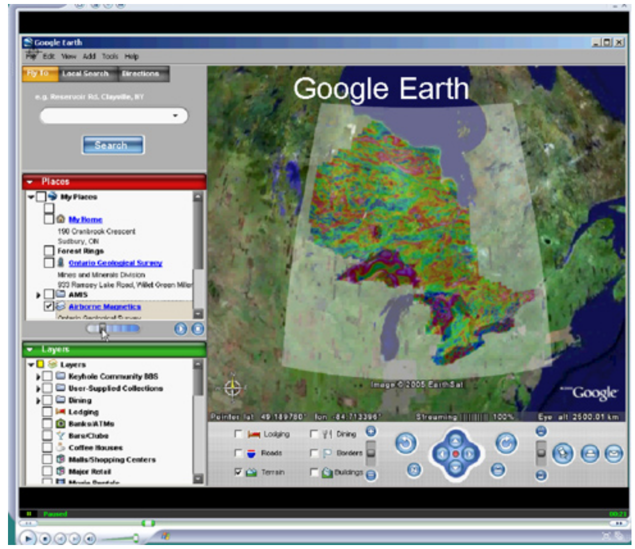


Figure 11. MNDM is reviewing best practices as far as incorporating ***Geology Ontario*** geoscience data into sites like Google Earth.

Building a Water Well Database for GIS Analysis

By A. Wayne Jones and Kelly A. Barrett

Ohio Department of Natural Resources
Division of Water
2045 Morse Rd. B-2
Columbus, Ohio 43229-6693
Telephone: (614) 265-1075
Fax: (614) 265-6767
e-mail: wayne.jones@dnr.state.oh.us

OBJECTIVES

The Ohio Department of Natural Resources (ODNR), Division of Water (DOW), is the official repository for all water well records in Ohio. Currently, over 775,000 paper well records are on file at the DOW (Figure 1). The water well records dataset has multiple uses, including requests for homeowner private well information, a usable resource for the water well drilling community, an information source for environmental consultants, and a base for mapping and ground water related research. The DOW is in the process of converting paper well logs to digital data in a database and scanning images of each well record. These data are accessed through a public database. To use these data for more advanced projects, like DOW modeling and potentiometric mapping, it is essential for the data type to be compatible with use of Geographic Information Systems (GIS). Ultimately, the database needs to support the public access activities, while also handling the more advanced GIS functions.

VENDOR

The DOW had three vendors for the data conversion and database construction project. The DOW staff and inmates at an Ohio state penal institution converted the paper well log data into a database. As these labor-intensive steps are costly, the DOW chose to work through a vendor who trained and supervised the inmates on data entry. A separate vendor scanned the well logs as images. Finally, a third vendor redesigned the "front-end" for the database that is used for public data requests. Staff members were responsible for overseeing these operations. A major challenge was quality assurance and quality control.

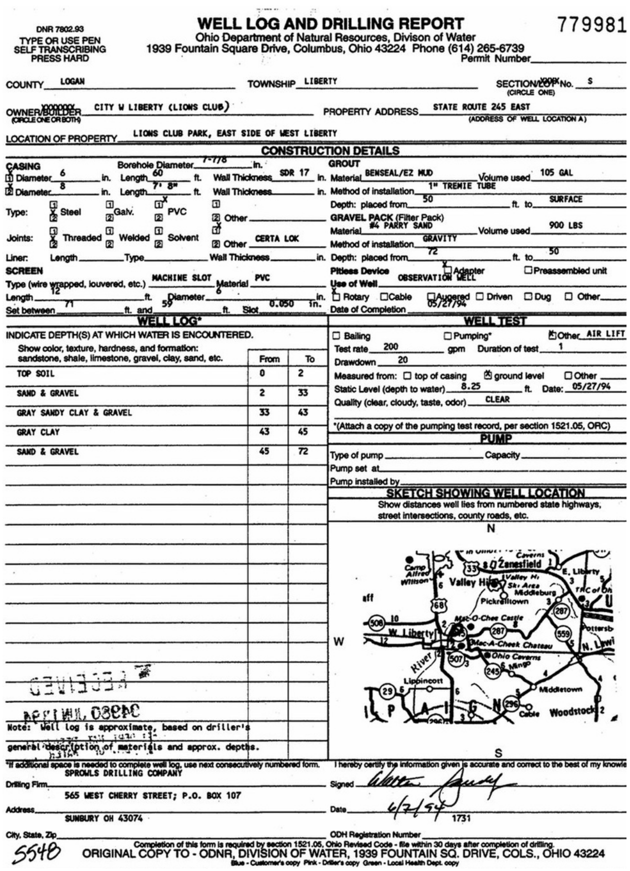
SOFTWARE USED

The well log database is an Oracle 10G Release 1-enterprise database. Oracle, which uses a Unix interface, is administered by the ODNR, Office of Information Technology. As an enterprise system, the DOW has limited write-authority and relies on the administrator to perform updates and maintenance. The DOW purchased from a vendor a custom-built Oracle front-end application called Flotiva from Workiviti version (8-1-6), which is used for data entry, simple string queries, and public service requests (Figures 2 and 3). The Oracle database is accessible to some internal users by using a link through Microsoft Access (2002). No write privileges are associated with this application. Queries and data dumps in Access are saved as tables that are usable in Microsoft Excel (2002). GIS applications are based on the Environmental Systems Research Institute (ESRI) products of Arc Map, Arc Catalog, Arc Toolbox (9.1) (Spatial Analyst, 3-D Analyst), and Arc Workstation (9.1).

WELL LOGS

The DOW began collecting water well records in 1947. Of the 775,000 paper well logs on file, less than 200,000 are located by spatial coordinates. Most of the well log locations are from DOW staff traveling into the field with a topographic map and the well log to plot the location. The field location process was abandoned in the early 1990s due to personnel reductions. Recently, geocoding has been used, with excellent results, for unlocated well log data for three counties.

Water well location maps (Figure 4) were digitized by the vendor and written to CD. The information was organized as an ESRI shapefile for each county in Ohio. To add the spatial coordinates of the wells to the Oracle well log database, a common field between the CD records and the database had to be identified. Each well log has a unique well log number, and therefore, this would have been the logical link between the digitized file and the Oracle database. Unfortunately, the well log locations were digitized using four fields (none of which are the well log number) to tag the locations. These four fields include county code, township code, location map year (the year in which the well logs were located in the field, since many counties



DNR 7802.93
TYPE OR USE PEN
SELF TRANSCRIBING
PRESS HARD

WELL LOG AND DRILLING REPORT 779981
Ohio Department of Natural Resources, Division of Water
1939 Fountain Square Drive, Columbus, Ohio 43224 Phone (614) 265-6739
Permit Number

COUNTY LOGAN TOWNSHIP LIBERTY SECTION/LOT No. 5
OWNER/BUILDER CITY W LIBERTY (LIONS CLUB) PROPERTY ADDRESS STATE ROUTE 245 EAST (ADDRESS OF WELL LOCATION A)
LOCATION OF PROPERTY LIONS CLUB PARK, EAST SIDE OF WEST LIBERTY

CONSTRUCTION DETAILS

CASING: Borehole Diameter 7-7/8 in. Diameter 6 in. Length 60 ft. Wall Thickness SDR 17 in. Diameter 8 in. Length 7' 8" ft. Type: Steel, Galv., PVC, Other. Joints: Threaded, Welded, Solvent, Other CERTA LOK. Liner: Length, Type, Wall Thickness.
SCREEN: Type (wire wrapped, louvered, etc.) MACHINE SLOT Material PVC Length 12 ft. Diameter 6 in. Set between 71 ft. and 59 ft. Slot 0.050 in.

GROUT: Material BENSEAL/EZ MUD Volume used 105 GAL. Method of installation 1" TREMIE TUBE. Depth: placed from 50 ft. to SURFACE.
GRAVEL PACK (Filter Pack): Material #4 PARRY SAND Volume used 900 LBS. Method of installation GRAVITY. Depth: placed from 72 ft. to 50.
Pitless Device: Adapter, Preassembled unit. Use of Well OBSERVATION WELL. Rotary, Cable, Augered, Driven, Dug, Other. Date of Completion 05/27/94

WELL LOG*

INDICATE DEPTH(S) AT WHICH WATER IS ENCOUNTERED.

Show color, texture, hardness, and formation: sandstone, shale, limestone, gravel, clay, sand, etc.	From	To
TOP SOIL	0	2
SAND & GRAVEL	2	33
GRAY SANDY CLAY & GRAVEL	33	43
GRAY CLAY	43	45
SAND & GRAVEL	45	72

Note: Well log is approximate, based on driller's general description of materials and approx. depths.
*If additional space is needed to complete well log, use next consecutively numbered form.

WELL TEST

Bailing, Pumping*, Other AIR LIFT. Test rate 200 gpm. Duration of test 1. Drawdown 20. Measured from: top of casing, ground level, Other. Static Level (depth to water) 8.25 ft. Date: 05/27/94. Quality (clear, cloudy, taste, odor) CLEAR.
*(Attach a copy of the pumping test record, per section 1521.05, ORC)

PUMP

Type of pump, Capacity, Pump set at, Pump installed by

SKETCH SHOWING WELL LOCATION

Show distances well lies from numbered state highways, street intersections, county roads, etc.

Drilling Firm SPROWLS DRILLING COMPANY
Address 565 WEST CHERRY STREET; P.O. BOX 107
City, State, Zip SUNBURY OH 43074

I hereby certify the information given is accurate and correct to the best of my knowledge.
Signed
Date 6/2/94
ODH Registration Number 1731

Completion of this form is required by section 1521.05, Ohio Revised Code - file within 30 days after completion of drilling.
ORIGINAL COPY TO - ODNR, DIVISION OF WATER, 1939 FOUNTAIN SQ. DRIVE, COLS., OHIO 43224
Blue - Customer's copy Pink - Driller's copy Green - Local Health Dept. copy

Figure 1. A typical water well log and drilling report.

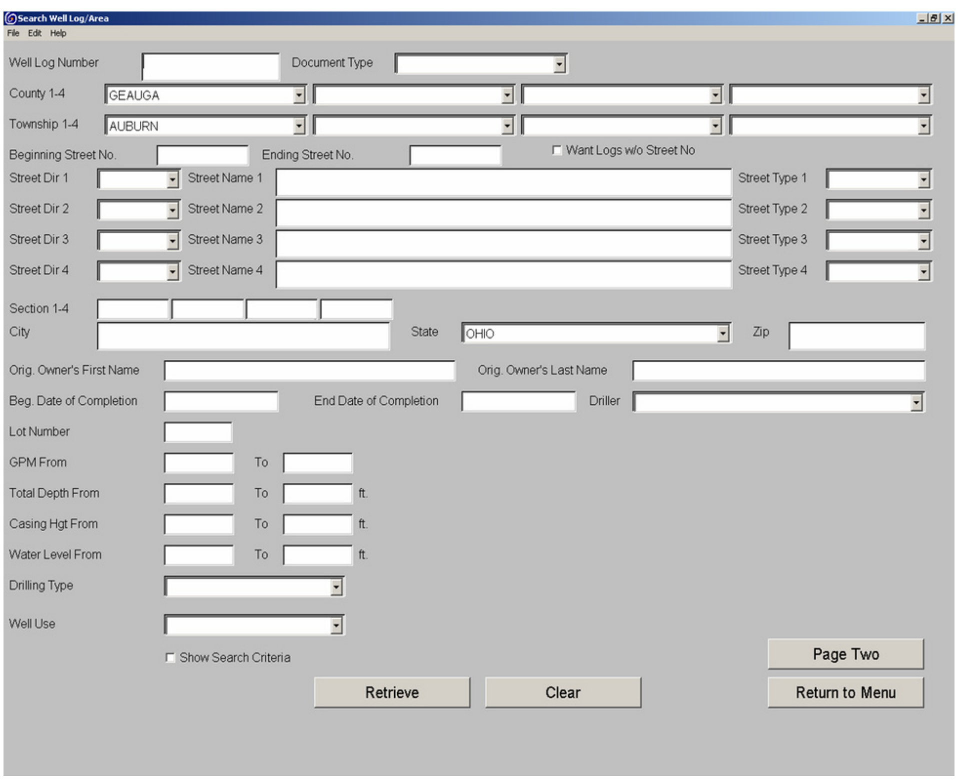


Figure 2. Search well log area screen, from Flotiva Oracle front-end custom application.

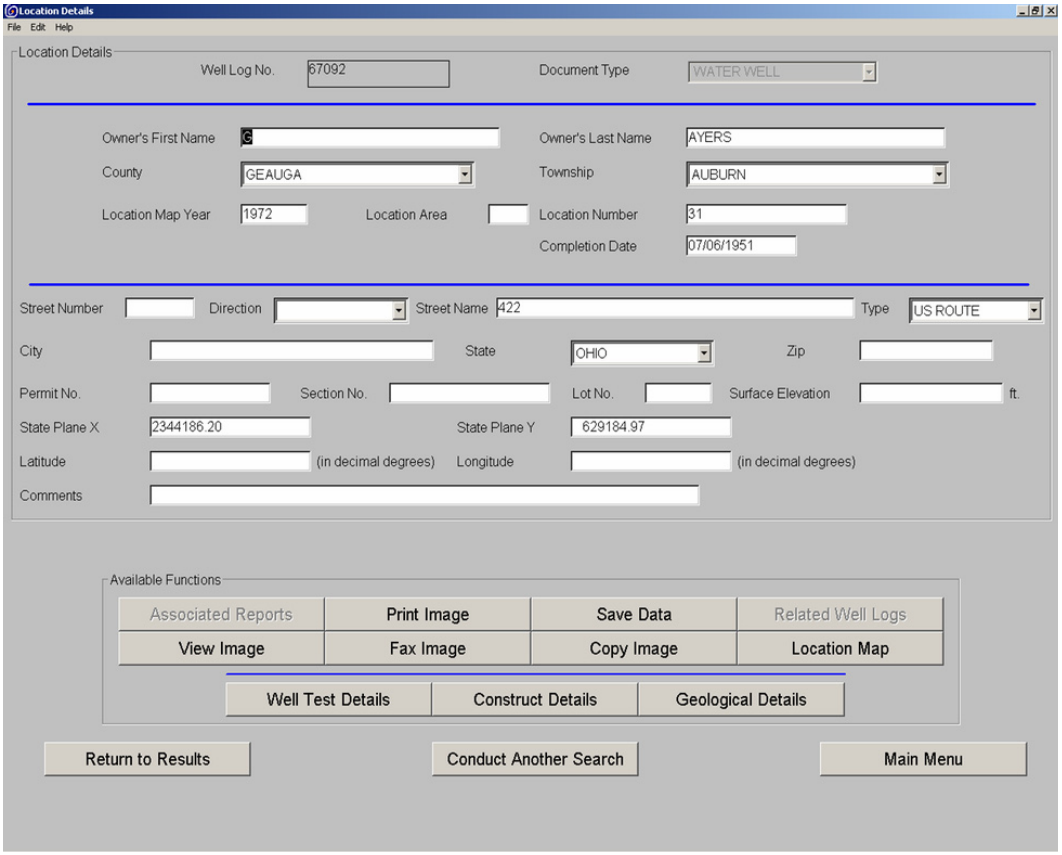


Figure 3. Location details entry screen, from Flotiva Oracle front-end custom application.

Figure 4. Map showing field-located sites of water well records.

have been field located multiple times), and location area. If these four fields match, then the link is made between the field location map points with spatial coordinates and the data fields in the Oracle database. In addition, paper copies of the well logs were scanned as TIFF type 4 images so they could be linked to the Oracle database.

Errors in linking CD data to the appropriate records in the database can be categorized as either duplicate errors or no-match (Figures 5 and 6). The shapefile of the digitized well locations is sent to the Oracle administrator to be run in a simulation to compare the digitized locations with the records in the Oracle database. The duplicate list shows all of the points with identical coordinates. Often, the errors come from inadvertently clicking the digitizing puck twice on a location. The no-match list contains those points where the four fields of data from the digitized file do not match the records in the well log database. No-match errors are either data entry or missing record errors in the Oracle database.

After error correction is completed, the Oracle administrator performs a finalized update to the Oracle database. Now, the well locations are linked to the database and to the scanned image of the well log.

```
geauga_dup.txt - Notepad
File Edit Format View Help
 >>>>  GEAUGA COUNT DUPLICATE LIST   <<<<
 ---------------------------------------------------------------
CNT TWP  LOC      LOC_NO  AR       X            Y
--- ---- ---- ---------- -- ---------- ----------
55  130  1972        449     2306230.74  639131.56
55  130  1972        449     2304664.52  637825.78

55  2450 1977        162      2306647.5  668203.81
55  2450 1977        162     2306605.72  656305.38

55  570  1987        389     2304903.85  673182.38
55  570  1987        389     2306569.91  675471.54

55  2450 1977         20     2318771.61  669803.24
55  2450 1977         20     2306947.01  656466.52

55  570  1987         29     2311677.93  676989.39
55  570  1987         29     2307359.37  673777.64

55  570  1987         44     2311353.05  691409.06
55  570  1987         44     2307354.64  673905.39

55  570  1987         45     2311153.46  691171.25
55  570  1987         45      2307156.1  673692.47

55  570  1945        882     2307828.43  688569.33
55  570  1945        882     2325381.96  679143.11

55  570  1987        548     2317699.67  689777.52
55  570  1987        548     2316707.44  690827.25

55  570  1987        563     2317755.25  685020.34
55  570  1987        563     2311269.83  679307.41

55  570  1987        521     2324873.06   692417.4
55  570  1987        521     2317902.54   695366.3

55  570  1945        947     2314910.21  685436.34
55  570  1945        947     2320985.21  672516.82

55  570  1945        489     2313394.78  686101.36
55  570  1945        489     2312908.09  686115.51

55  90   1945       1249     2341258.95  629068.77
```

Figure 5. List of duplicates, showing matching records with different spatial coordinates.

POTENTIOMETRIC SURFACE MAPPING

Mapping ground water flow directions (potentiometric surface mapping) is extremely important to most hydrogeological studies. The DOW is currently producing potentiometric surface maps for Ohio. Producing electronic datasets to support potentiometric surface mapping from our database is a high priority.

The process of generating a potentiometric surface map begins by downloading the necessary fields from the database using Access (2002). Important fields are well log number, county, township, location map year, location number, static water level, and geologic formation. The data are complied in Excel (2002) and converted into a DBF4 table. The table is imported into ArcView 9.1 as an event theme, then converted into a shapefile. This shapefile is edited to confirm that the well's coordinates lie in the correct county and township (Figure 7).

The Digital Line Graphs (DLG) hypsography layer for each quadrangle is downloaded from the Center for Mapping at the Ohio State University (Figure 8). The DLGs are projected to State Plane South NAD27 coordinate system. The hypsography (elevation contours)

```
geauga_xynomatch.txt - Notepad
File Edit Format View Help
>> GEAUGA X AND Y NOMATCH LIST <<
-----------------------------------
Cty Twp Yr  Loc Ar  X         Y
--- --- --  --- --  -         -
55,130,1945,976,,2309170.93,621910.75,
55,130,1945,713,,2308134.78,620102.02,
55,130,1945,1216,,2306534.99,620186.5,
55,130,1945,598,,2310947.05,616490.49,
55,130,1972,14,,2322237.05,619460.01,
55,130,1972,294,,2322905.62,620895.38,
55,130,1972,107,,2322513.5,615982.62,
55,130,1972,108,,2322813.98,616005.18,
55,130,1972,310,,2323305.31,618536.82,
55,130,1972,277,,2325304.27,621900.76,
55,130,1972,196,,2322882.34,622453.51,
55,130,1972,191,,2319637.12,622631.72,
55,130,1972,192,,2319697.21,622849.77,
55,130,1972,195,,2320448.42,622646.76,
55,130,1972,190,,2320718.86,622646.76,
55,130,1972,131,,2312824.01,615983.8,
55,90,1972,144,,2334646.4,615928.99,
55,130,1945,1165,,2320219.55,622647.03,
55,130,1945,949,,2319693.3,623027.19,
55,130,1945,1050,,2318556.6,618613.1,
55,130,1945,1040,,2323393.88,622778.08,
55,90,1945,515,,2334152.5,620112.21,
55,130,1972,382,,2309502.05,627750.26,
55,130,1945,488,,2310158.61,625809.99,
55,130,1972,263,,2327045.58,627199.35,
55,130,1972,211,,2325501.41,629588.3,
55,130,1972,276,,2327729.1,631876.49,
55,130,1972,476,,2318259.35,630254.31,
55,130,1972,498,,2316255.15,628387.35,
55,130,1972,473,,2313413.85,634410.9,
55,130,1945,908,,2319055.57,633666.61,
55,130,1945,569,,2312348.28,633907.16,
55,130,1945,864,,2322636.17,628292.32,
55,130,1945,972,,2322867.63,623284.96,
55,90,1945,445,,2329887.78,633034.24,
55,90,1945,1282,,2333796.75,629427.81,
55,130,1972,456,,2305277.39,639159.96,
55,130,1972,457,,2306236.42,638433.52,
55,130,1972,349,,2304738.29,638155.21,
55,130,1945,37,,2304752.64,641118.49,
55,2450,1988,119,,2315432.79,645066.07,
```

Figure 6. The no-match list (well log database records where the actual record is lost).

layer is recoded to display the "elevation" item using code written in arc macro language (AML). The resulting surficial elevation from the hypsography layer is used to construct a Triangular Integrated Network (TIN) model (Figure 9). A TIN model is a 3-dimensional surface of the X, Y, and Z values interpolated at locations between data points. The ArcInfo Workstation command Tinspot selects the elevation at the spatial location of each point in the coverage and picks the elevation from the TIN model. This elevation item will be significantly more accurate than selecting the elevation from a DEM surface, where the cell size can skew the value returned. The static water level is subtracted from the surface elevation at each point to get the elevation of the water table. USGS 1:24,000

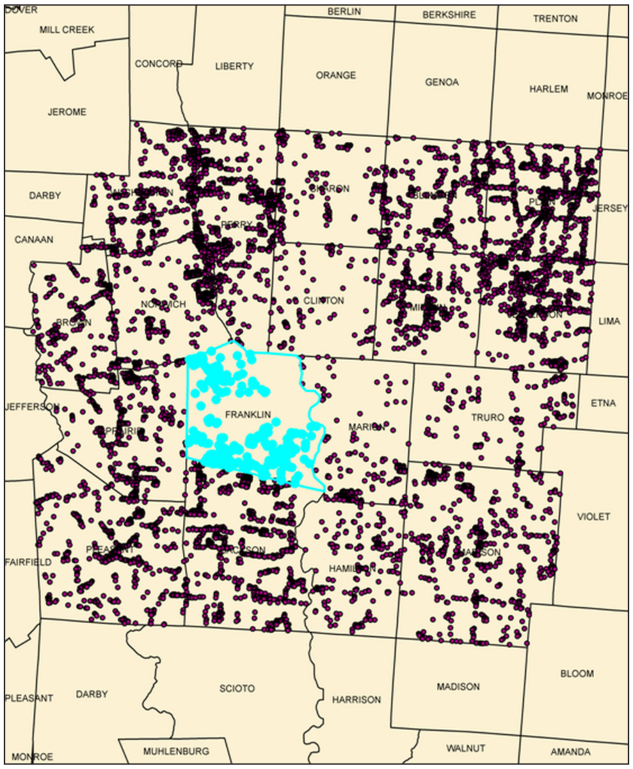


Figure 7. Water well location map. Franklin County wells highlighted in light gray (light blue in online version).

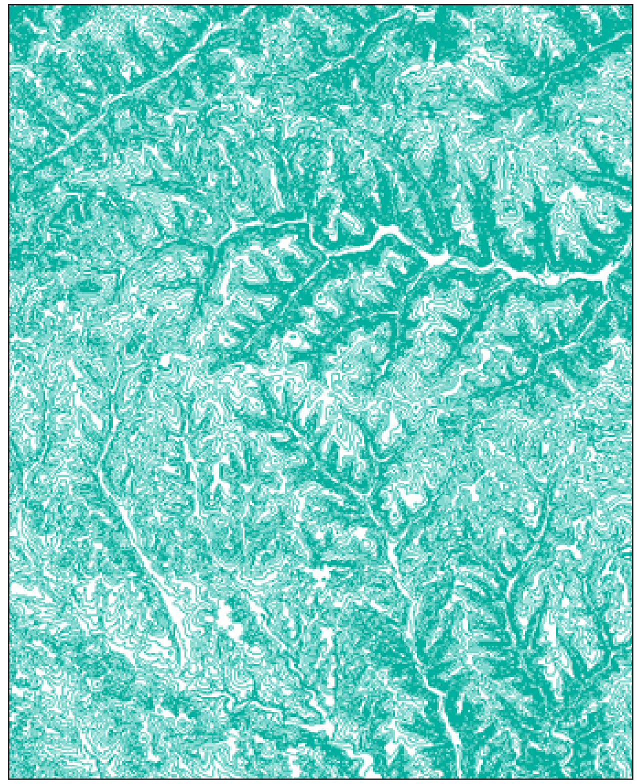


Figure 8. A typical Digital Line Graph (digitized topographic elevation contours).

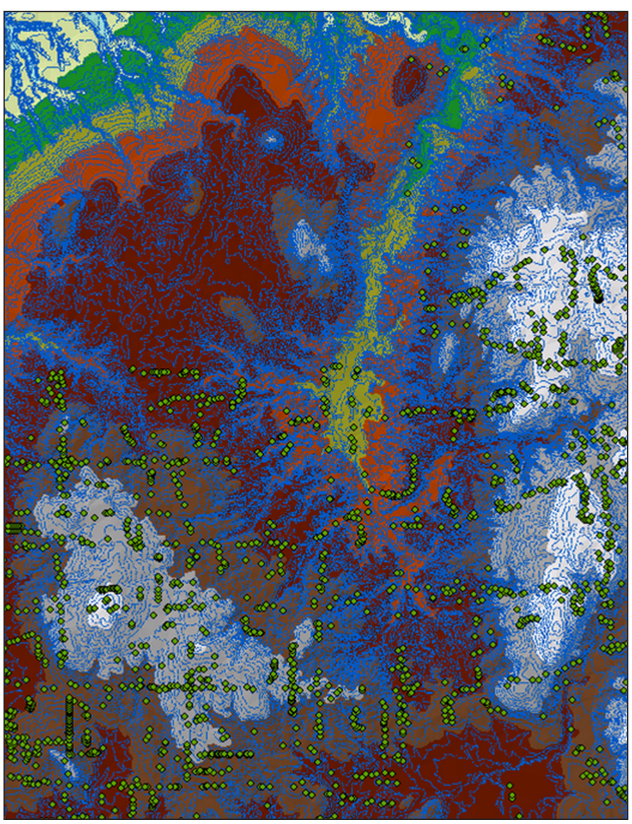


Figure 9. The 3-D TIN (Triangular Irregular Network) model constructed from Digital Line Graph file, then shaded by elevation. Well log points also are shown.

quadrangle plots of the water elevation point data are printed on Mylar for mapping hydrogeologists to interpret and map ground water flow direction. After mapping is completed, the contours drawn on the Mylar sheets are digitized to create a final layout (Figure 10).

USING GIS TO CREATE AND ANALYZE POTENTIOMETRIC SURFACE

In our latest effort with the digital database, the data points from the well log database were plotted to create sand and gravel and a bedrock aquifer potentiometric surface map near Darby Creek in Western Franklin County, Ohio. These potentiometric surface maps were combined to show where groundwater is flowing. The shapefiles were turned into TIN models, which were turned into grids. Using grid subtraction in ArcMap, Spatial Analyst, a raster difference map was generated from the potentiometric surface map data. On the Difference Map, negative values (blues) show areas where water can move downward from the sand and gravel to the bedrock aquifers. Positive values (reds) show areas where the bedrock aquifer has a net upward effect, which means that water moves from the bedrock into the sand and gravel. For more information on this technique, please view the PDF of this work as a poster session from the DMT 06 conference, http://ngmdb.usgs.gov/Info/dmt/docs/angle06.pdf.

APPLICATIONS FOR DIGITAL DATA

The DOW receives many requests for our digital water well log database. The requests may be for an area, county, watershed, or localized site. Using the capacities of GIS to clip the data to the desired configuration has made completing those requests easier. Some requests (often from environmental consulting companies) are for an area within a given radius from a natural feature or

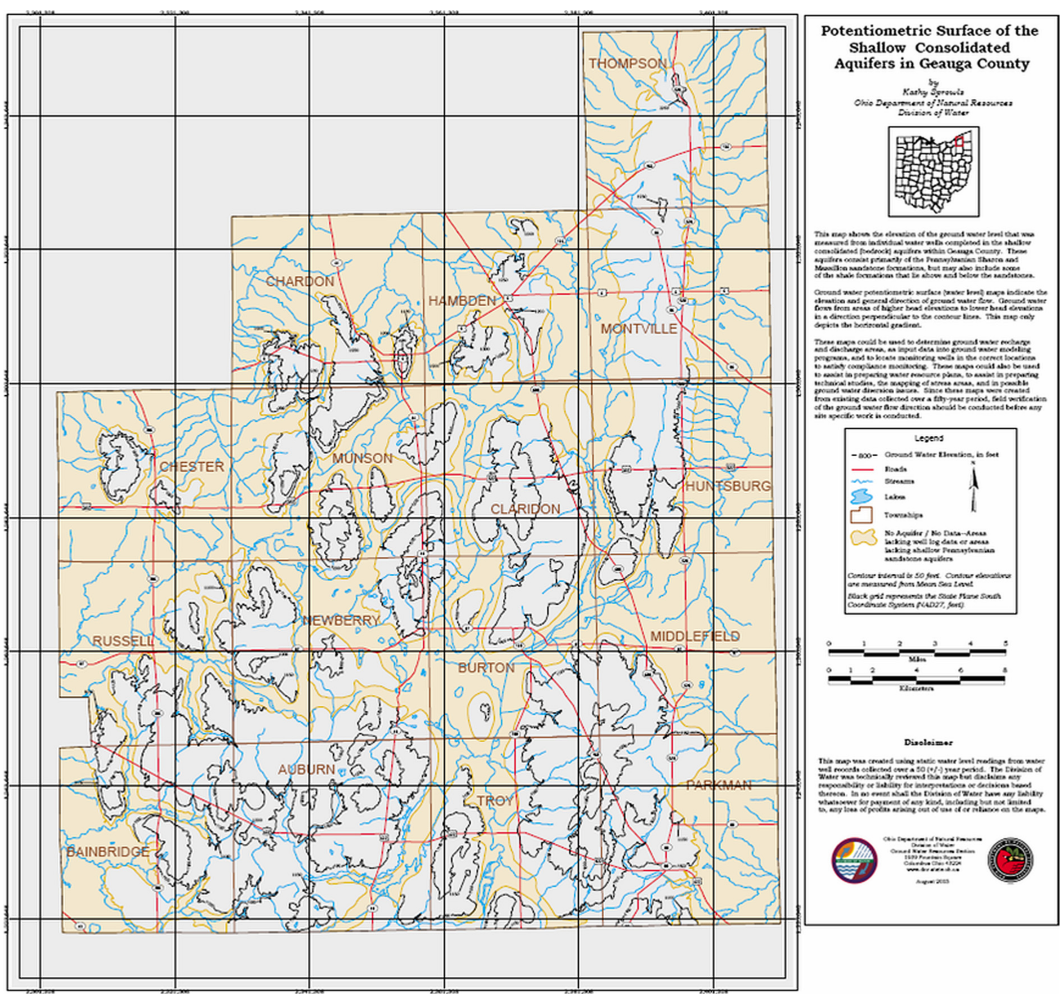


Figure 10. Potentiometric surface map (contour map of water table elevation) of shallow consolidated aquifers in Geauga County, Ohio.

intersection. The buffer function in ArcMap makes these requests straightforward (Figure 11). For example, if you were looking for ground water/surface water interaction, you can easily create a query that returns all of the water wells within a 1-mile radius of a major stream. A further refinement of the query might be to return only the sand and gravel wells with a total depth of less than 50 feet and a static water level of less that 20 feet. Consulting firms often request a 1-mile radius around a site (Figure 12). GIS capacity allows for quick fulfillment of these requests.

FINAL PRODUCTS FOR DIGITAL DATA

Templates are developed in ArcMap for all of our mapping projects. This allows for a standard design to all maps of the same series. The templates require the input of new data and some minor modifications and the map is published on-line as a PDF. Examples of these maps can be viewed on-line at http://www.dnr.state.oh.us/water/gwpsurface/County_List/tabid/3621/Default.aspx and http://www.dnr.state.oh.us/water/gwppmaps/default/tabid/3541/Default.aspx.

SOFTWARE CITED

ESRI, ARCGIS – Environmental Systems Research Institute, Inc., 380 New York St., Redlands, CA 92373-8100 USA, (909) 793-2853, http://www.esri.com

Flotiva-Workiviti, Document Imaging Solutions, 8529 N. Dixie Drive, Dayton, Ohio 45414 USA, 1-937-890-5135, http://www.disolutions.com

Microsoft, Microsoft Office 2002 – Microsoft Corporation, One Microsoft Way, Redmond, WA 98052-6399 USA, 1-800-642-7676, http://www.microsoft.com

Oracle, Oracle 10G Release 1 – Oracle Corporate Headquarters 500 Oracle Parkway, Redwood Shores, CA 94065 USA, 1-650-506-0024, http://www.oracle.com

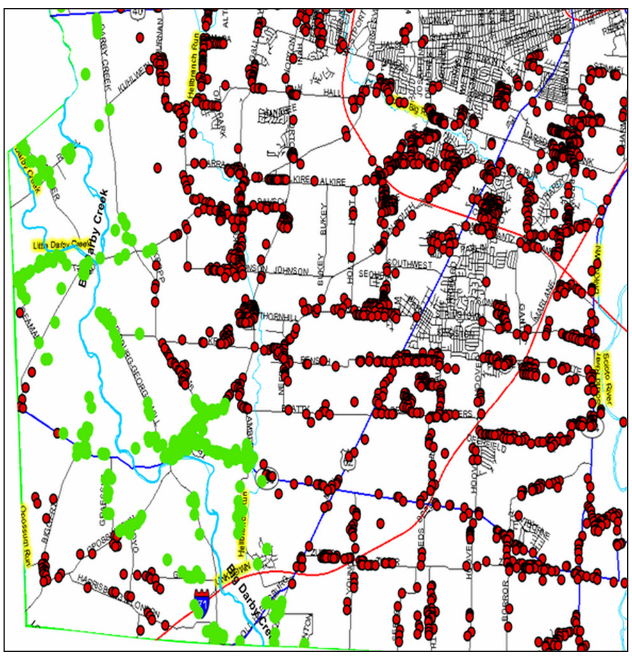


Figure 11. Water wells within one-mile buffer around Big Darby Creek in Franklin County, Ohio. Wells within buffer are highlighted in light gray (green in online version).

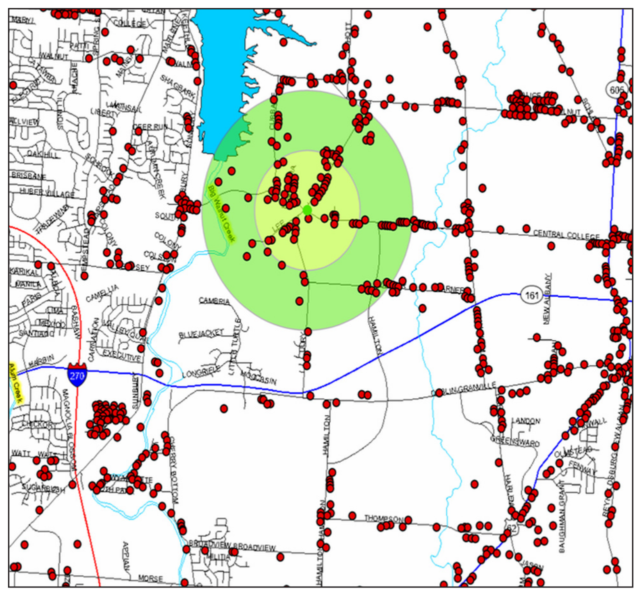


Figure 12. One-mile search radius centered on a road intersection, indicating water wells within the search area.

Sensitive Aquifers in Ohio – Relationship to Highly Susceptible Public Water Systems

By Christopher Kenah, Michael W. Slattery, Linda D. Slattery, and Michael L. Eggert

Ohio Environmental Protection Agency
Division of Drinking and Ground Waters
50 W. Town Street
Columbus, OH 43215
Telephone: (614) 644-2752
Fax: (614) 644-2909
e-mail: christopher.kenah@epa.state.oh.us

INTRODUCTION

One of Ohio Environmental Protection Agency's goals is to identify aquifers in Ohio that are sensitive to ground water contamination as a result of land use activities. The primary benefit of identifying sensitive aquifers within the state is to help prioritize limited resources in order to maximize ground water protection efforts. This article outlines the approach Ohio has used to identify sensitive aquifers (Ohio EPA, 2006) by integrating geologic and chemical water quality data with digital mapping to develop a derivative map product. Data utilized in this effort includes geologic information presented in Ohio Department of Natural Resources Aquifer Maps (ODNR, 2000), water quality information from Ohio's Ambient Ground Water Monitoring Program, and Ohio's public water system (PWS) water quality compliance monitoring data. PWSs with chemical water quality impacts (criteria outlined in Ohio's Source Water Protection and Assessment Program) are also used to evaluate Ohio's sensitive aquifers. The complex nature of ground water contamination will always require site-specific investigations to identify sources and pathways for impacts to ground water. These site-specific studies help refine our hydrogeologic knowledge and may result in refinement of sensitive aquifers. The goal here, however, is to use statewide data to identify aquifer settings that are most likely to be impacted by land use activities.

OHIO SENSITIVE AQUIFERS

The major aquifers in Ohio include widespread, unconsolidated sand and gravel units, sandstone bedrock in the eastern half of Ohio, and carbonate bedrock in the western half, as illustrated in Figure 1. The sand and gravel aquifers fill pre-glacial and glacial valleys cut into bedrock and are referred to as buried valley aquifers. The sandstone and carbonate bedrock aquifers generally provide sufficient production for water wells except where dominated by shale, as in southwest and southeast Ohio. Glacial drift overlies most of Ohio except for the unglaciated southeastern quarter of the state.

An understanding of ground water recharge pathways and water quality data were used to identify those aquifers most likely to be impacted by land use activities; that is, sensitive aquifers. The concept that short or rapid recharge pathways increase aquifer sensitivity is widely accepted and used to identify and evaluate sensitive aquifers. Applying this concept within the state of Ohio suggests that sand and gravel aquifers are the most sensitive. Shallow bedrock aquifers, particularly fractured or karst bedrock aquifers that underlie thin glacial drift (tills or lacustrine deposits), comprise a second group of sensitive aquifers. Elevated nitrate concentrations from PWS compliance monitoring data confirmed that these aquifers were sensitive based on ground water quality impacts (Ohio EPA, 2003a). An underlying assumption is that the distribution of potential contaminant sources is widespread and evenly distributed, and the results of this analysis reflect the influence of recharge pathways, not potential source distribution. The widespread distribution of nitrate sources (agricultural, residential) in Ohio makes this a reasonable assumption; however, high concentrations of potential sources close to a PWS well increases the likelihood that the well will exhibit water quality impacts.

The geologic information used in this analysis was derived from the Glacial Aquifer Map (ODNR, 2000), and consequently, the analysis does not provide information about the sensitivity of aquifers not included within, or occurring below, glacial deposits. Geologic settings with rapid recharge are identified as sensitive aquifers and include the following as described in the ODNR Glacial Aquifer Map:

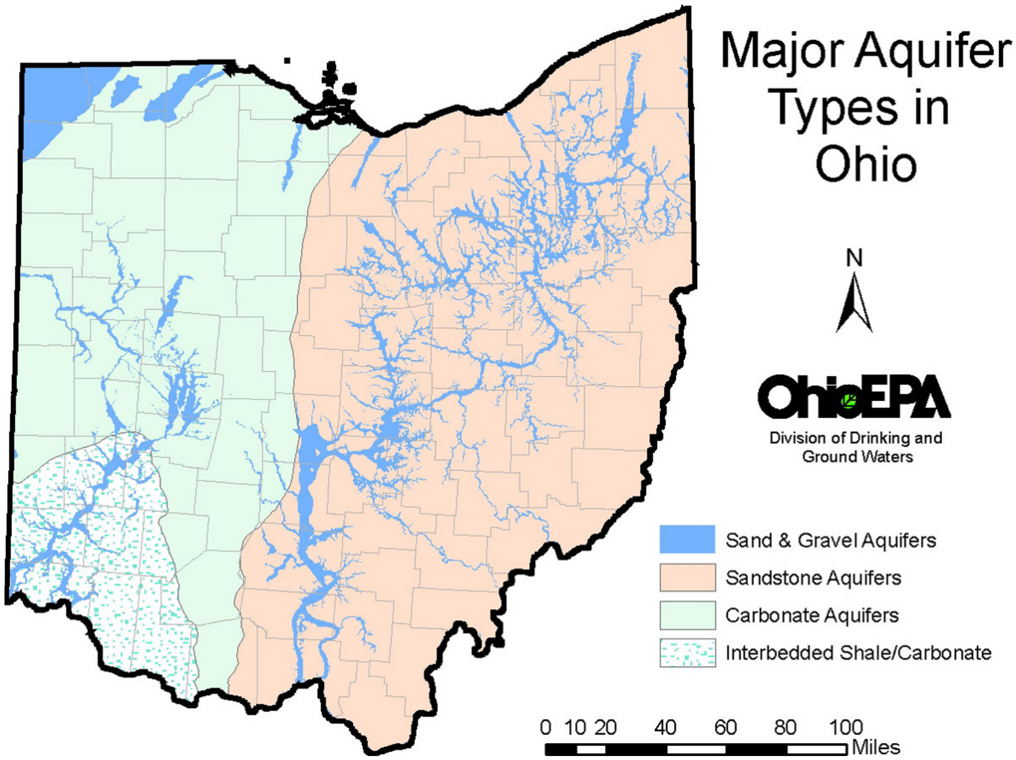


Figure 1. Distribution of major aquifers in Ohio, modified from ODNR Aquifer Maps (Ohio EPA, 2000).

Sand and Gravel Aquifers:
- Buried Valley
- Alluvial
- Valley Fill
- Outwash/Kame
- Beach Ridge

Bedrock Aquifers Below Thin Uplands and Lacustrine Deposits (<25 feet):
- Thin Uplands (thin till)
- Lacustrine

In contrast, areas of thick till (e.g., glacial moraines) generally retard recharge, thereby reducing the sensitivity of aquifers within or below the tills. The following geologic settings in the ODNR Glacial Aquifer Map are not sensitive where the till is relatively thick (>25 feet):

Moraine Deposits:
- Ground Moraine
- End Moraine
- Complex

These geologic settings are associated with different glacial material thicknesses across the state. It is generally assumed that the greater the glacial thickness, the greater the protection provided to the aquifer. The longer recharge pathways and increased recharge travel time to the aquifer reduces the overall sensitivity of the aquifer. The ODNR Aquifer Maps groups glacial drift into three thickness categories: thin (<25 feet), moderate (25-100 feet), and thick (>100 feet). This thickness describes either the thickness of the glacial drift that includes the aquifer or the thickness of glacial material that overlies a bedrock aquifer. Each of the hydrogeologic settings was separated into these thickness groups to evaluate whether water quality impacts were influenced by the glacial overburden thickness. Even though these groupings are coarse, differences between geologic setting and thickness groups are clear, as discussed in later sections.

This analysis was also performed using lithologic attributes included in the Glacial Aquifer Maps. The lithologic attributes describe the primary materials within mapped polygons in the ODNR Aquifer Maps and provide further division of some geologic settings. The lithologic parameters were divided into three groups for this analysis:

- Sand and Gravel Lithologies: Includes coarse to fine sand and gravel units with minor fine-grained material, including thin lenses of alluvium, lacustrine deposits, or till.
- Fine Grained Lithologies: Predominantly fine grained geologic materials with minor sand and

gravel lenses. Alluvium, slack water, till, and colluvium deposits are included in this group.
- Till Lithologies: Predominantly tills with little evidence of sand and gravel lenses. In thin tills, wells generally penetrate through the till into bedrock aquifers. In thicker tills, limited production may be associated with sand and gravel lenses, but larger production wells will generally be drilled through the till into bedrock aquifers.

Overall, the coarser lithologies would be expected to allow more rapid recharge and, consequently, be associated with more sensitive aquifers. Figure 2 illustrates the distribution of the sand and gravel aquifers that are sensitive, and the areas of thin glacial drift that overlie sensitive bedrock aquifers.

PWSs WITH GROUND WATER QUALITY IMPACTS

The 1996 Amendments to the Safe Drinking Water Act established a program for states to assess drinking water sources for all public water systems (PWSs). The purpose was to provide PWSs information for developing drinking water protection plans. These assessments included a susceptibility analysis. As outlined in the Source Water Assessment and Protection (SWAP) Program, Susceptibility Analysis Process Manual (Ohio EPA, 2003b), a PWS in Ohio is considered to have "high susceptibility" if it has been impacted by anthropogenic contaminants, regardless of its geologic setting or lithology. Water quality impacts are defined as two or more nitrate concentrations greater than 2.0 mg/L, or two or more confirmed detections of organic constituents (VOC or SVOC) using PWS compliance monitoring data since 1991. The SWAP staff reviewed more than 60,000 samples, including over 1,100,000 results, to identify the subset of high susceptibility PWSs based on water quality impacts. This effort identified a subset of 561 PWSs out of a total of 5,151 ground water sourced PWSs. This subset of PWSs can be used as an independent data set to evaluate the identification of sensitive aquifers in Ohio. Figure 3 shows the locations of the 561 PWSs with documented water quality impacts (highly susceptible in SWAP terminology) in relationship to sensitive aquifers. The visual association of the PWSs that exhibit water quality impacts to sensitive aquifers is most obvious along buried valleys, but is not particularly obvious on a state scale map.

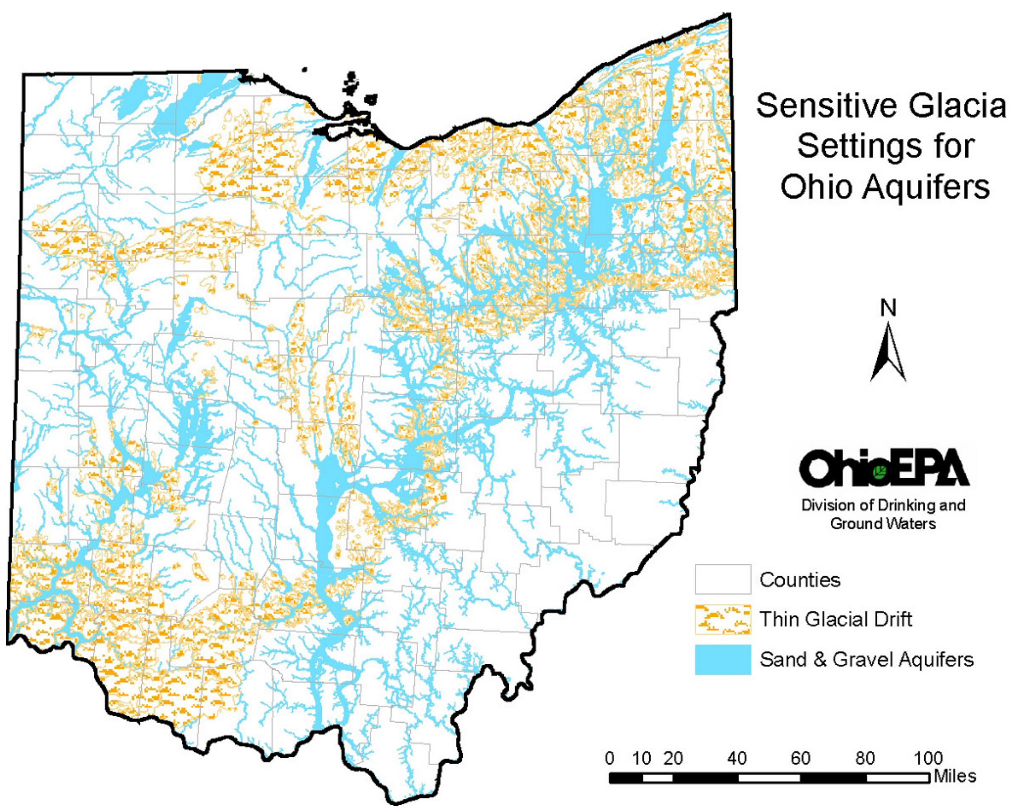


Figure 2. Distribution of sensitive sand and gravel aquifers and thin glacial drift over sensitive bedrock aquifers.

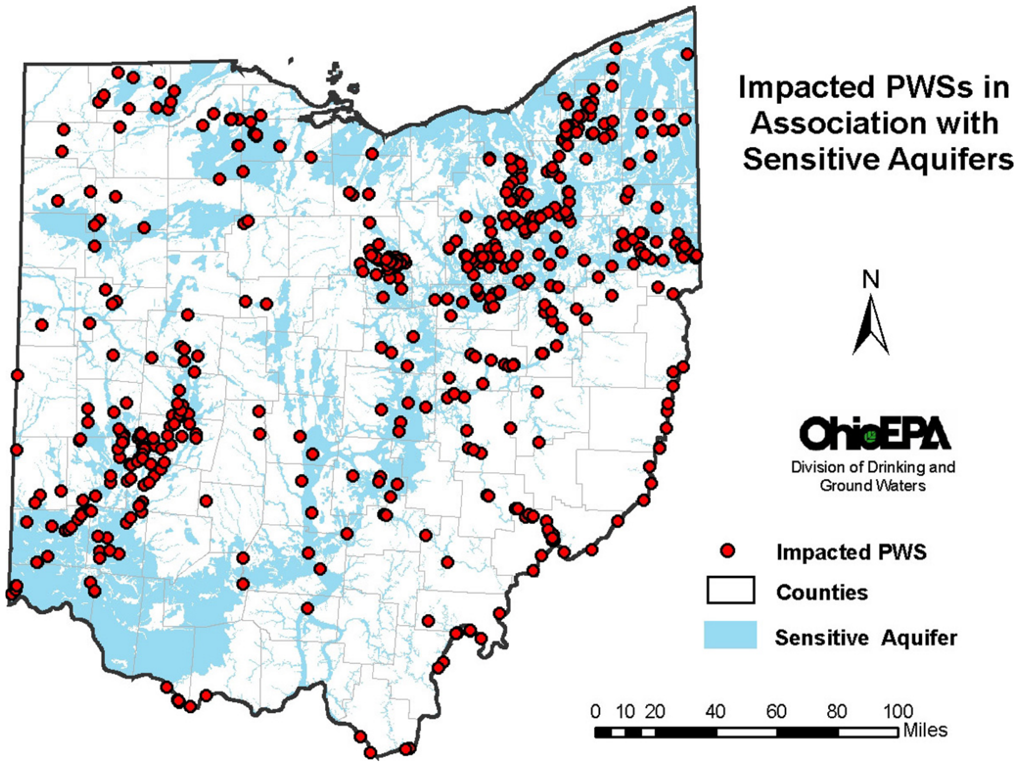


Figure 3. PWSs with documented water quality impacts in relation to sensitive aquifers.

ASSOCIATION OF HIGH SUSCEPTIBILITY PWSs WITH SENSITIVE AQUIFERS

The following analysis utilized the ODNR Glacial Aquifer maps to determine whether the highly susceptible PWSs are located where sensitive aquifers occur, with the implication that the well is probably completed in that aquifer. To accomplish this, the 561 highly susceptible PWSs were associated by location to the attributes of the ODNR Glacial Aquifer Maps. These attributes, including geologic setting, thickness, and lithology, were used to count the number of impacted PWSs that occur in different glacial settings or the number associated with different glacial lithologies. These counts were compared to the number of all ground water based PWSs associated with the same groupings of glacial settings or lithologies to normalize the results for accurate comparisons. The following section presents the results as percentages of ground water-based PWSs.

Extracting the Glacial Aquifer Map's attribute data based on PWS locations has several limitations that need to be considered with regard to the results presented in Tables 1 and 2. The thickness of the glacial drift at a given PWS location is not the well depth or average well depth of the PWS wells, but rather an estimate of the glacial drift thickness that controls the recharge pathways. The PWS well may be producing water from the glacial drift, or the well may be cased through the glacial drift and producing water from a bedrock aquifer. If the glacial thickness is thin (< 25 feet thick), it is reasonable to deduce that the well is a bedrock well. In thicker glacial drift, it is not possible to know whether the well is producing water from the glacial aquifers or bedrock aquifers without additional information such as well depth or casing length. If the PWS wells are located in thick glacial drift in geologic settings that are considered sensitive (buried valley, beach ridge, outwash/kame), it is likely that the well is producing from sand and gravel aquifers that lie within the drift. For example, a well may be 45 feet deep in an area of drift greater than 100 feet thick. These data limitations need to be taken into account when reviewing the data presented in Tables 1 and 2. Overall, there appears to be a strong association between impacted PWSs and glacial attributes associated with sensitive aquifers, in spite of these data limitations.

GEOLOGIC SETTING ASSOCIATIONS

The percentages of impacted PWSs associated with categories of geologic settings by increasing thickness of glacial drift are listed in Table 1. Overall the highest percentages of PWSs with impacted source water are associated with the sand and gravel aquifers. The major aquifer groups associated with geologic settings are discussed below.

Table 1. Documented water quality impacts at PWSs, and associations between impacts and hydrogeologic setting* of glacial units.

Thickness of Glacial Unit (Feet)	Number of ground water-based PWSs with Impacts	Total Number of ground water-based PWSs	Percentage of Impacted PWSs
Sand and Gravel Aquifer Settings			
(Aquifer Map settings - Buried Valley, Alluvial, Valley Fill, Outwash/Kame, Beach Ridge)			
0-25	10	39	**26%**
25-100	92	487	**19%**
>100	153	1066	**14%**
Bedrock Aquifers Below Thin Uplands and Lacustrine Deposits			
(Aquifer Map settings - Thin Uplands and Lacustrine)			
0-25	133	904	**15%**
25-100	103	1032	**10%**
>100	8	28	**29%**
Moraine Deposits			
(Aquifer Map Settings - Ground Moraine, End Moraine, Complex)			
0-25	0	0	**0%**
25-100	27	704	**4%**
>100	13	559	**2%**
Unglaciated Areas			
(No glacial units on ODNR Aquifer Map)			
0	22	332	**7%**

*Hydrogeologic settings are from ODNR Glacial Aquifer Map (ODNR, 2000).

Table 2. Documented water quality impacts at PWSs, and associations between impacts and lithology* of glacial units.

Thickness of Glacial Unit (Feet)	Number of ground water-based PWSs with Impacts	Total Number of ground water-based PWSs	Percentage of Impacted PWSs
Sand and Gravel Lithologies			
(including: sand and gravel, minor fines, confined, thin till included within or over unit)			
0-25	6	9	**67%**
25-100	50	133	**38%**
>100	113	495	**23%**
Fine Grained Lithologies			
(including: fine grained sediments undifferentiated, fines with minor sand and gravel lenses)			
0-25	4	36	**11%**
25-100	44	405	**11%**
>100	37	464	**8%**
Till Lithologies			
(including: till, till with sand and gravel lenses)			
0-25	133	898	**15%**
25-100	128	1682	**8%**
>100	24	697	**3%**
Unglaciated Areas			
(no glacial units on ODNR Aquifer Map)			
0	22	332	**7%**

*Lithology divisions are from ODNR Glacial Aquifer Map (ODNR, 2000).

Sand and Gravel Settings

The percentage of PWSs with documented water quality impacts is higher for the geologic settings dominated by sand and gravel deposits (buried valley, alluvial, valley fill, outwash/kame, and beach ridge) as illustrated in Table 1, column 4. The percentage of impacted systems decreases with increasing glacial drift thickness: 26 percent for thin sand and gravel deposits (statistic based on small number of PWSs), 19 percent for intermediate sand and gravel deposits, and 14 percent for sand and gravel deposits of more than 100 feet thick. This trend supports the concept of increased protection as the recharge pathways lengthen. Based on the statutory requirement of 25 feet of casing for PWS wells, wells associated with the 0-25 foot group of sand and gravels probably represent wells drilled through the glacial sand and gravel into bedrock.

Bedrock Aquifers Below Thin Uplands and Lacustrine Deposits

PWS wells producing from bedrock aquifers below thin and intermediate-thickness upland till or lacustrine deposits (< 25 and 25-100 feet thick) exhibit ground water quality impacts in 15 and 10 percent of the PWSs, respectively. These percentages suggest that thin till and lacustrine material provides a bit more protection from land use activity than sand and gravel. The low yield from till and lacustrine deposits requires most of these wells to be bedrock production wells. The 29 percent of PWSs that showed ground water impacts with glacial drift greater than 100 feet thick is based on relatively few wells and appears anomalous. Two of the eight impacted wells include sensitive buried valley or beach ridge geologic settings within their Drinking Water Source Protection Area inner management zone (one year time-of-travel). These settings provide rapid recharge pathways. The other six wells are deep wells (>120 feet) and, based on available well logs, are presumed to produce water from confined aquifers. This is an unlikely set of wells to exhibit elevated nitrate, since most nitrate detections > 2.0 mg/L occur in Ohio wells with depths less than 75 feet. Three of these four wells were drilled in the 1950's. One explanation is that well construction deficiencies or corrosion of the casing may allow leakage of shallow, nitrogen rich ground water into these wells. Several of these wells are within a few miles of each other and exhibit similar nitrogen time series patterns, which suggests a regional rather than a local control. This example illustrates some of the problems with applying broad interpretations, as shown in Table 1, to site-specific cases.

Moraine Deposits

Protection of ground water resources by thick till cover is demonstrated in the moraine deposits category of Table 1, where the percentage of impacted PWSs does not exceed 4 percent. In Table 1, the thin tills are grouped with the thin upland and lacustrine deposits; 15 percent of PWSs located in thin till settings exhibit water quality impacts. Thicker tills, however, are associated with relatively few impacted PWSs; 4 percent for the intermediate (25-100 feet) group, and 2 percent for the thick (>100 feet) group of the moraine deposits category. These low percentages suggest that thick till provides significantly more protection than sand and gravel or thin tills. The role that fractures and macropores play in controlling recharge to aquifers through tills is a current topic of discussion in Ohio (Weatherington-Rice and Christy, 2000; Weatherington-Rice and others, 2006). Fractures and macropores certainly affect the movement of recharge through thin tills; this is supported by the high percentage (15 percent) of impacted PWSs associated with thin upland and lacustrine settings. Fractures and macropores appear to be significantly less important in recharge and contaminant transport in thick tills as documented by the much lower percentages (4 percent and 2 percent) of impacted PWSs associated with thicker moraine deposits. This is believed to be caused by the limited vertical extent of fractures in till, typically 20-25 feet (Scott Brockman, personal communication, 2006).

Unglaciated Areas

The last category in Table 1 includes PWSs located in the unglaciated areas of Ohio, primarily the southeastern uplands. These areas include weathered bedrock (colluvium) that overlies late Paleozoic sandstones and shales, and in places is overlain by loess deposits. In most cases, wells produce water from bedrock aquifers; however, yields are generally low and relatively few PWSs use these aquifers. Consequently, we have limited data for the unglaciated areas. The available data, however, indicates that the colluvium is better than sand and gravel and thin till in glaciated areas, but not as good as thicker tills at protecting wells from land use impacts. From Table 1, PWSs in unglaciated areas are impacted in 7 percent of the sites, a value that is intermediate between sand and gravel and thicker tills. The lower population density and reduced number of potential pollution sources in unglaciated, southeastern Ohio uplands also tends to depress the percentage of impacted PWSs in this region.

LITHOLOGIC ASSOCIATIONS

The lithologic attributes can be divided for further analysis into sand and gravel, fine grained, and till lithologies. The percentages of impacted PWSs associated with these lithologic groups, ordered by increasing thickness of glacial drift, are presented in Table 2. The percentage of PWSs with documented water quality impacts is highest for the sand and gravel lithologies and lowest for the till materials. This is similar to the results presented in Table 1 for geologic settings, but the percentages associated with the lithologic divisions indicate that recharge processes are more dependent on the lithologic material than on the geologic setting. The major lithologic groups are discussed individually below.

Sand and Gravel Lithologies

For sand and gravel lithologies, the low number of wells in the < 25 feet thickness category results from the statutory requirement that PWS wells have a minimum of 25 feet of casing installed. Thus, the few wells in this thickness grouping are either old wells with short casing lengths or bedrock wells cased through thin glacial sand and gravel. With 67 percent of these wells having water quality impacts, it appears that they are highly susceptible, but because the percentage is based on a very low number of wells, it cannot be given much significance. The percentage of impacted PWSs decreases to 38 percent and 23 percent for the 25-100 feet and >100 feet groupings, respectively. These wells probably do not penetrate the entire glacial thickness, but it appears that the thicker section of sand and gravel provides more filtration and increased travel time, as one would expect.

Almost all of the areas of sand and gravel lithologies in Table 2 are included within the geologic settings that are considered most sensitive in Table 1. The sensitive geologic settings that include sand and gravel aquifers also include fine grained lithologies. The higher correlation between coarser sand and gravel lithologies and PWSs with water quality impacts (Table 2) emphasizes the importance of identifying coarser sand and gravel lithologies as more sensitive than finer grained sand and gravel lithologies. This is consistent with the understanding that aquifer sensitivity is controlled by recharge and transport rate of surface or near surface contaminants to aquifers.

Fine Grained Lithologies

The fine grained portions of the glacial lithologies in the ODNR Aquifer Maps are predominantly undifferentiated fine grained sediments with minor sand and gravel lenses. Approximately 8 percent to 11 percent of the PWSs associated with these fine grained lithologies exhibit water quality impacts, which is significantly less than coarse sand and gravel deposits (23 percent to 38 percent for the sand and gravel lithologies). The percentage of impacted PWSs decreases with increasing glacial drift thickness, but not appreciably, which may result from the variability of well depths independent of the glacial deposit thickness. The low number of samples in the 0-25 feet thickness group (36) suggests that the result (11 percent) should not be given much significance, except that the 25-100 feet group exhibits the same percentage of impacted PWSs.

Till Lithologies

Table 2 documents a sizable difference between impacted PWSs in thin tills and the thicker tills. Only limited protection is provided by thin glacial drift overlying bedrock aquifers, as evidenced by the 15 percent of impacted PWSs associated with thin tills (< 25 feet). Wells producing from fractured bedrock just below thin glacial cover will be the most vulnerable and probably account for the bulk of the PWSs with water quality impacts in this category. The percentage of PWSs with water quality impacts drops dramatically for thicker tills: 8 percent for tills in the 25-100 feet group, and 3 percent for tills greater than 100 feet thick.

Unglaciated Areas

The unglaciated areas include the same subset of wells in Table 2 as presented in Table 1. As stated earlier, the limited data restrict broad conclusions, but the low percentage of impacted PWSs suggests that colluvium provides some protection for these wells from land use activities. The lower population density and reduced number of potential pollution sources in southeast Ohio may also help to keep this percentage low.

DISCUSSION

This analysis documents the importance of distinguishing the coarser grained lithologies in buried valleys, alluvial, valley fill, outwash/kame, and beach ridge deposits as being more sensitive to contaminant impact than the finer grained deposits in these same settings. Figure 4 illustrates the association of the PWSs with water quality impacts along a section of the Great Miami buried valley aquifer in southwest Ohio. This figure illustrates

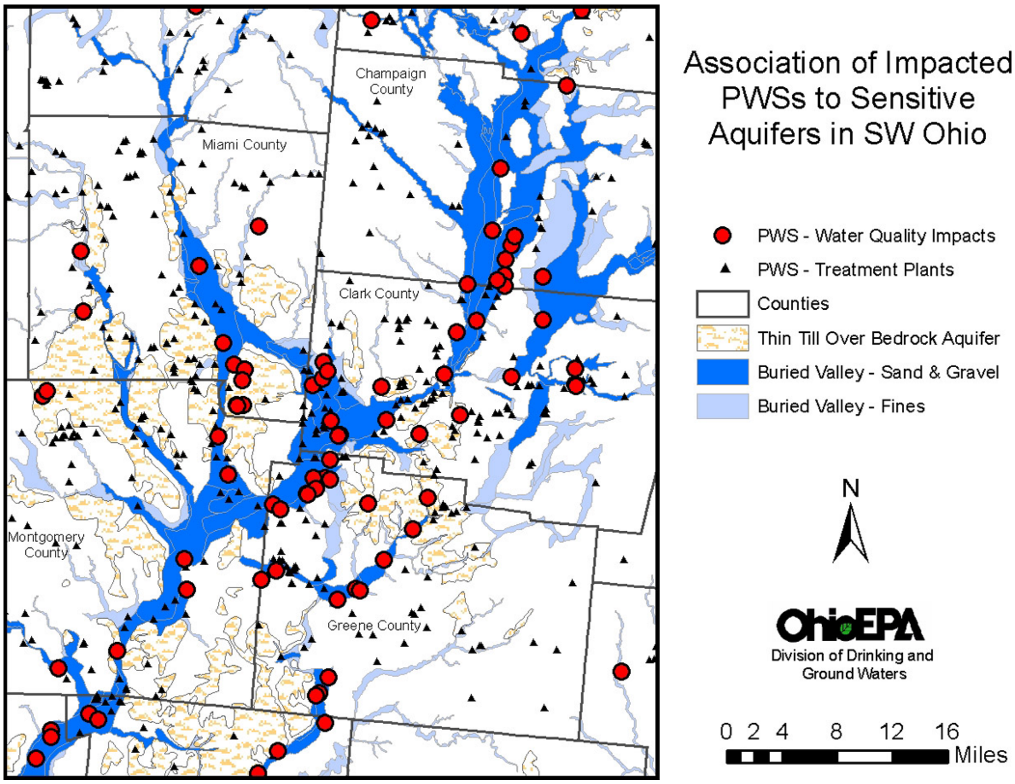


Figure 4. Impacted PWSs and sensitive aquifers and in southwest Ohio.

the relationship between the high concentrations of water quality impacts and sections of the buried valley identified as sand and gravel. Water quality impacts are also associated with finer grained buried valley deposits and thin tills in the uplands (where wells penetrate the till to produce from bedrock aquifers), but with lower frequency as documented in Table 2. The locations of PWS treatment plants are provided to illustrate the wide distribution of PWSs used in the analysis.

Highlighting Sand and Gravel Aquifers

The divisions on the glacial aquifer map are general groupings that were selected to represent glacial material on a state scale based on geologic and well log data. Analyses using PWS empirical water quality data (elevated nitrate, VOC detections) exhibit significant correlations with the identified sensitive geologic settings and lithologies. The stronger association of impacted PWSs with coarser and more permeable sand and gravel lithologies underscores the role recharge plays in making aquifers sensitive, and suggests that coarse sand and gravel units should be emphasized in the identification of Ohio's sensitive aquifers.

Figure 5 illustrates the distribution of sensitive glacial geologic settings across Ohio, utilizing attributes identified in ODNR's Glacial Aquifer Maps (ODNR, 2000). This figure represents the combined analysis of sensitive aquifers and water quality impacts based on nitrate and VOC concentrations. In the glacial settings where sand and gravel deposits are common, the lithology is divided into fine and coarse grained units. The analysis presented in this paper documents that the coarser grained units are more sensitive. The classification processes used in developing the glacial aquifer maps required a great deal of simplification, and consequently, these generalizations need to be considered in any application of Figure 5. Nevertheless, the correlation of the empirical PWS water quality data to the lithologic and geologic setting descriptors supports the simplifications made in developing the ODNR Glacial Aquifer map and the validity of using recharge controls in determining sensitive aquifers.

The goal of identifying sensitive aquifers is to help set priorities for protecting the state's ground water resources. Statewide, aquifers that have more than 25 feet of till or fine grained glacial deposits overlying the well production zone are less likely to exhibit anthropogenic water quality impacts than unconfined sand and gravel aquifers. If till that overlies a bedrock production aquifer is less than 25 feet thick, the bedrock aquifer's sensitivity is elevated, but the sand and gravel aquifers are even more sensitive. The data summaries in Tables 1 and 2, and Figure 5, illustrate the high geologic sensitivity (based on

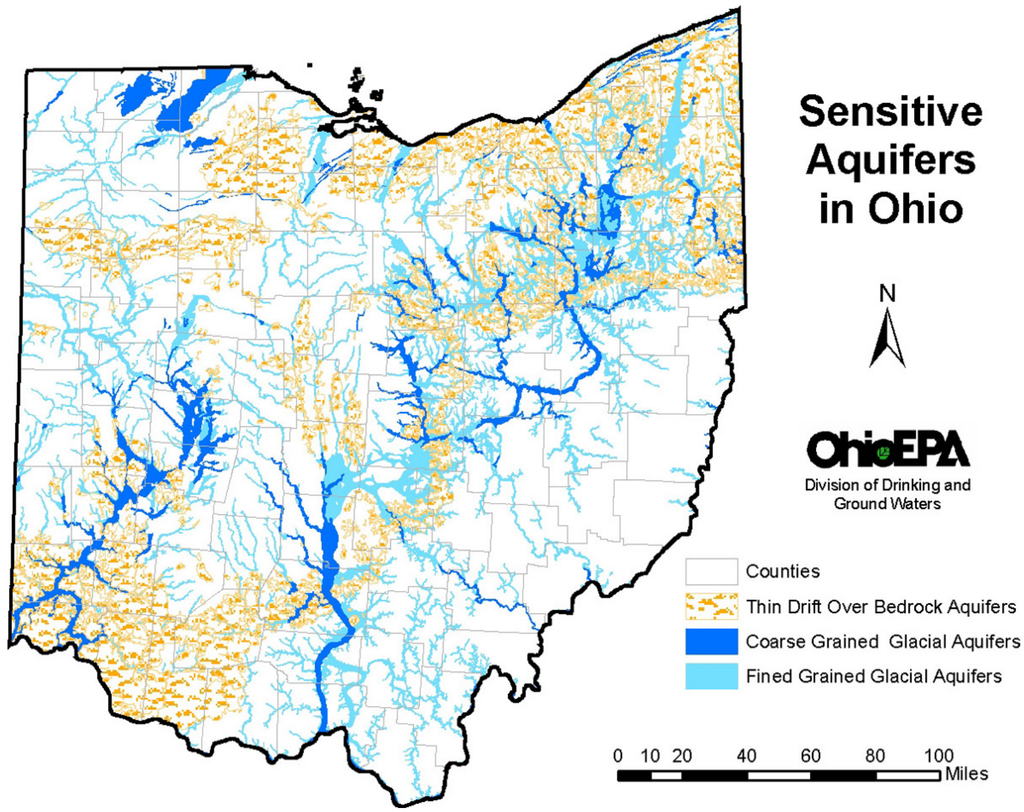


Figure 5. Distribution of sensitive aquifers in Ohio.

chemical water quality data) of Ohio's productive sand and gravel aquifers, in particular the areas of coarser sand and gravels. These conclusions are not unexpected, and they support the generally accepted views of hydrogeologists familiar with the fate and transport of contaminants within sensitive aquifers in Ohio. These empirical data provide significant support for judgments based on professional experience and, consequently, increase our confidence in applying the identified sensitive aquifer settings to protecting Ohio's ground water resources.

Thin Till Over Bedrock Aquifers

Several situations in areas of thin glacial till (< 25 feet) over bedrock have been identified and emphasize the sensitivity of this geological setting, especially where the till that overlies the bedrock is less than 10 feet. Wastewater management (septic, manure) and agricultural or residential facilities placed in close proximity to private or public wells can result in health issues if contaminants move rapidly through the thin till along macropores, fractures, or other pathways to fractured bedrock aquifers. It appears that microbiological contamination is more likely in these thin till settings than in sensitive, unconsolidated sands due to the limited filtration capacity of macropores and fractures in till and bedrock. Invariably, ground water microbiological contamination increases proportionally to increasing population concentration in areas of closely spaced septic systems and wells. Figure 6 is a schematic geologic cross section that illustrates the sensitive geologic setting of fractured bedrock below thin glacial drift. Water quality impacts associated with similar geologic setting are recognized across the state. The 2004 infectious disease outbreak at South Bass Island in Lake Erie, which was determined to be associated with ground water contamination (Ohio Department of Health, 2005), is an example of the potential water quality impact to sensitive bedrock aquifers in areas of thin to no glacial overburden.

Horizontal Flow Pathways

The previous analysis has assumed that recharge pathways are dominantly vertical. Although this is generally true, there are geologic settings where recharge contributions can have significant horizontal components that increase the sensitivity of aquifers. The schematic cross section in Figure 7 illustrates this situation. Wells located in flood plains close to rivers are examples of locations where horizontal flow paths can become significant contributors to recharge. Many of these wells are located close to the river to increase production by inducing recharge from the river. At flood stage, the elevated river

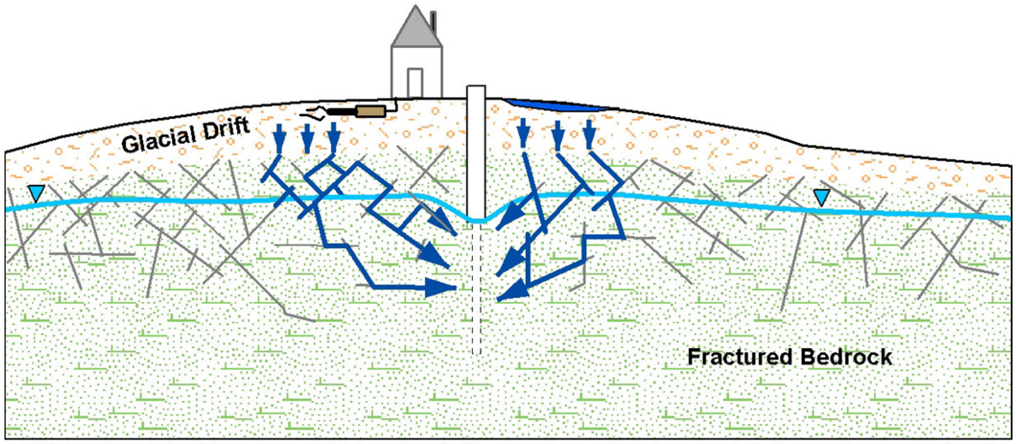


Figure 6. Schematic cross section illustrating rapid pathways in areas of thin glacial drift. The ground water table is indicated by the thick line with triangles above it.

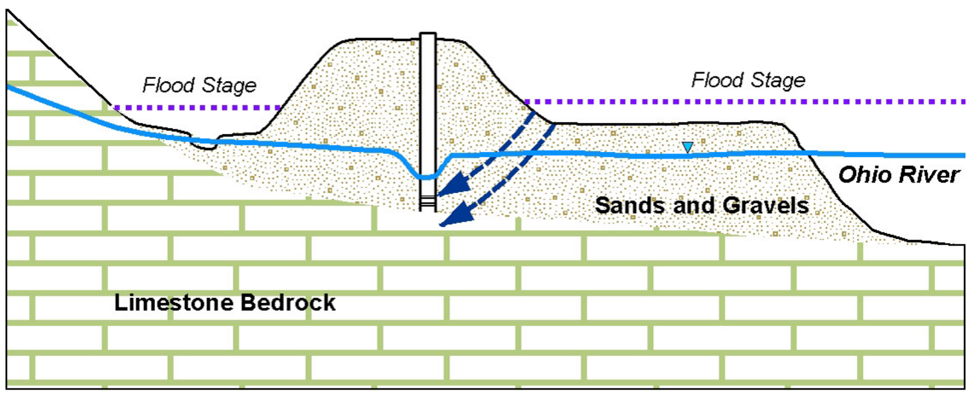


Figure 7. Schematic cross section illustrating shortened flow paths for wells in flood plains.

level can shorten the horizontal flow path and increase the pressure gradient to a pumping well, with the result of shortening recharge pathways and reducing transport time. There are also areas where fractured bedrock exposed in streams and stream banks can provide rapid recharge to bedrock aquifers, especially during flooding events. The areas of the state where horizontal components of recharge are potentially significant also need to be identified as areas of sensitive aquifers.

CONCLUSIONS AND FUTURE ACTIVITIES

This approach to identifying sensitive aquifers demonstrates a practical integration of digital geologic mapping and water quality data analysis to generate a derivative map of sensitive aquifers. The intention of this map is to help prioritize ground water protection activities. The map is certainly not a final product but rather a work in progress. The current draft illustrates the core logic of using recharge pathways to determine sensitivity and documents the validity of the approach based on water quality impacts. Refinements of this map will include:

- Assignment of glacial drift thickness to PWS wells for more precise analysis of water quality impact and drift thickness;
- Incorporation of well depth and casing length into the analysis;
- Evaluation of differences in aquifer sensitivity between dissolved components and particulate components such as pathogens;
- Completion of analysis of variability of aquifer sensitivity as related to increased concentration or more persistent occurrences of anthropogenic contaminants; and
- Identification of areas where horizontal flow paths are significant.

REFERENCES

Ohio Department of Health, 2005, Director Investigation - Gastrointestinal Illness, South Bass Island, Lake Erie, August 2004: Multiple Agency Investigation and Actions, February 22, 2005; amended February 16, 2006, accessed at http://www.epa.state.oh.us/ddagw/SBIweb/Reports/SBI_ODH.pdf.

Ohio Department of Natural Resources, 2000, Glacial Aquifer Map and Bedrock Aquifer Map (digital format), accessed at http://www.dnr.state.oh.us/water/samp/.

Ohio EPA, 2000, Ohio's Ground Water Quality, 2000 305(b) Report: Division of Drinking and Ground Waters, 38 p., accessed at http://www.epa.state.oh.us/ddagw/pdu/gw_305b.html.

Ohio EPA, 2003a, Ohio's Ground Water Quality, 2002 305(b) Report: Division of Drinking and Ground Waters, 78 p., accessed at http://www.epa.state.oh.us/ddagw/pdu/gw_305b.html.

Ohio EPA, 2003b, Susceptibility Analysis Process Manual: Division of Drinking and Ground Waters, 64 p., accessed at http://www.epa.state.oh.us/ddagw/Documents/swap_saprocman.pdf.

Ohio EPA, 2006, Ohio's Ground Water Quality, 2006 305(b) Report: Division of Drinking and Ground Waters, 93 p., accessed at http://www.epa.state.oh.us/ddagw/pdu/gw_305b.html.

Weatherington-Rice, J., and Christy, A.D., Eds. 2000, Fractures in Ohio's Glacial Tills: The Ohio Journal of Science, Special Issue. Vol. 100, no.3/4, accessed at http://www.oardc.ohio-state.edu/fractures.

Weatherington-Rice, J., Christy, A. D., and Angle, M.P., Eds., 2006, Fractures in Ohio's Glacial Tills - Further Explorations: The Ohio Journal of Science, Special Issue, Vol. 106, no.2, assessed at http://www.oardc.ohio-state.edu/fractures/.

GeoSciML – A GML Application for Geoscience Information Interchange

By Stephen M. Richard and CGI Interoperability Working Group

U.S. Geological Survey and Arizona Geological Survey
416 W. Congress #100
Tucson, AZ 85701
Telephone: (520) 770-3500
Fax: (520) 770-3505
e-mail: steve.richard@azgs.az.gov

CGI Interoperability Working Group is (in alphabetic order): Eric Boisvert, Boyan Brodaric, Simon Cox, Tim Duffy, Jonas Holmberg, Bruce Johnson, John Laxton, Tomas Lindberg, Stephen Richard, Alistair Ritchie, Francois Robida, Marcus Sen, Jean Jacques Serrano, Bruce Simons, and Lesley Wyborn.

INTRODUCTION

The GeoSciML application is a standards-based data format that provides a framework for application-neutral encoding of geoscience thematic data and related spatial data. GeoSciML is based on Geography Markup Language (GML, Cox et al., 2004) for representation of features and geometry, and the Open Geospatial Consortium (OGC) Observations and Measurements Best Practices (Cox, 2006) for observational data. Geoscience-specific aspects of the schema are based on a conceptual model for geoscience concepts and include geologic unit, geologic structure, and Earth material from the North America Data Model (NADMC1, North American Geologic-Map Data Model Steering Committee, 2004), and borehole information from the eXploration and Mining Markup Language (XMML, https://www.seegrid.csiro.au/twiki/bin/view/Xmml/WebHome). Development of controlled vocabulary resources for specifying content to realize semantic data interoperability is underway.

The intended scope for initial versions of GeoSciML includes information typically found on geologic maps as well as information typically recorded with boreholes. The possible uses for GeoSciML include transporting, storing, and archiving information. Amongst these, the most significant is transport—or information exchange—which enables information to be visualized, queried, and downloaded in spatial data infrastructures. This role for GeoSciML is particularly important, as geoscience information consumers are becoming more digitally sophisticated and are no longer satisfied with images and portrayals of data, but want digital data in standardized formats that can be used immediately in applications. Hours, days, or weeks spent merging data sets obtained separately from multiple agencies is time wasted. Use of a standardized markup for serializing geoscience information supports a commitment by data providers to publish data to users in a standardized format. Thus, GeoSciML allows applications to utilize globally distributed geoscience data and information.

The GeoSciML (https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/GeoSciML) project was initiated in 2003 under the auspices of the Commission for the Management and Application of Geoscience Information (CGI) working group on Data Model Collaboration (https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/WebHome). The CGI is a commission of the International Union of Geological Sciences and has the objective to enable the global exchange of geoscience information for legal, social, environmental, and geoscientific reasons. The project is part of what is now known as the CGI Interoperability Working Group (https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/InteroperabilityWG), which has the specific objectives to:

- develop a conceptual model of geoscientific information that draws on existing data models,
- implement an agreed subset of this model in an agreed schema language,
- implement an XML/GML encoding of the model subset,
- develop a test bed to illustrate the potential of the data model for interchange, and
- identify areas that require standardized classifications to enable interchange.

GeoSciML draws from many geoscience data model efforts and from them establishes a common suite of feature types based on geological criteria (units, structures, fossils) or artifacts of geological investigations (specimens, sections, measurements). Supporting objects are also considered (timescale, lexicons, etc), so that they can be used as classifiers for the primary objects. Predecessor projects that have had a strong influence on the development of GeoSciML include activities undertaken within national statutory bodies (e.g., the USGS/AASG National Geologic Map Database, British Geological Survey, and Japanese Geological Survey) in multi-jurisdictional contexts (the North American Data Model, http://nadm-geo.org/, for geological maps), and activities oriented to an industry sector (eXploration and Mining Markup Language – XMML, https://www.seegrid.csiro.au/twiki/bin/view/Xmml/WebHome). Currently, several external projects are leveraging GeoSciML for more specific applications, including Water Resources monitoring and management, Soils, Geotechnical and Engineering, Assay Data, and Geochemistry.

This report summarizes the schema and instance documents as implemented in a test bed demonstrated at the IAMG meeting in Liege, Belgium in September, 2006. The working group met subsequent to the test bed demonstration and has identified a number of aspects of the model and schema in need of update, as well as model elements that need to be added. Anticipated changes are discussed here as well. Version 1.1 is the current version of the markup language, with schema available at https://www.seegrid.csiro.au/subversion/xmml/GeoSciML/tags/1.1.0/schema/. Planning is underway for evolution of the schema to version 2 to expand the scope and clarify some of the top level model issues. Working group activity is currently focused in several task-groups (pending formalization):

- Use-cases and requirements task group, responsible for setting technical goals.
- Design task group, responsible for the structural and syntactic aspects of the "Information Model" of a GeoSciML-based service architecture.
- Service architecture task group, responsible for the "Computational Model" of GeoSciML-based service architecture.
- Concepts definition task group, responsible for the "Semantic Model," which will be a standard set of concepts (ontology) for the content used to populate GeoSciML, and will facilitate semantic interoperability with GeoSciML.
- Implementation test bed task group, responsible for liaison with GeoSciML Design and Service Architecture task groups to ensure that requirements are satisfied and coordinate and deliver TestBed3 demonstrating the GeoSciML v.2 use-cases.
- Outreach and technical assistance task group, responsible for providing advice and assistance to direct collaborators, assisting them to deploy conformant GeoSciML services.

GEOLOGIC MAP DATA SCHEMATIC INTEROPERABILITY

The development of standardized markup languages is a critical step necessary to achieve interoperability, which is defined by ISO/IEC 2382-01 (SC36 Secretariat, 2003) to mean: "The capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units." Technical requirements to meet this goal include system-level shared protocols for network communication, resource discovery, and service invocation (Figure 1). Applications that use these protocols must communicate by way of a shared data language that defines how information will be encoded. Geography Markup Language

Figure 1. Multiple levels of interoperability (Brodaric and Gahegan, 2006).

(http://www.opengeospatial.org/standards/gml) is the data language adopted for GeoSciML development. GML provides a framework for encoding geometry, defining features and associating them with properties (including geometry), and constructing dictionaries in which controlled vocabularies can be defined.

GeoSciML is a GML application scheme, which is defined by a collection of XML schema that utilize and extend elements from GML to represent standard geologic observations and descriptions in a geospatial context. GeoSciML is not a database structure. GeoSciML defines a format for data interchange (Figure 2). Agencies can provide a GeoSciML interface onto their existing data base systems, with no restructuring of internal databases required.

The semantic level of interoperability (Figure 1) requires agreement on the meaning of words used to express property values contained in GeoSciML elements. Developing common meanings for GeoSciML contents that can be applied to various multi-lingual vocabularies is a planned future activity. At present, we anticipate that implementation of schematic interoperability will demonstrate the need for data content standards to enable semantic interoperability.

IMPLEMENTATION

GeoSciML was developed by representatives from an international group of geologic map data providers in a series of face-to-face meetings and online discussion (see Twiki at https://www.seegrid.csiro.au/twiki/bin/view/CGI-Model/GeoSciML). One design objective was to re-use, revise, and extend existing standards wherever possible. The design philosophy of this interchange format has focused on an accurate representation of geoscience information in a general way. This results in great representational flexibility at the price of complexity and verbose encoding. Fortunately, text-based XML compresses very efficiently, and the markup is designed for machine input and output, not human readability.

Model development has utilized UML notation with a UML profile to enable systematic mapping from UML to XML schema. The mapping from UML models to GML is described in https://www.seegrid.csiro.au/twiki/bin/view/AppSchemas/UmlGml and https://www.seegrid.csiro.au/twiki/bin/view/AppSchemas/UmL2GMLAS. A detailed procedure for generating a GML-compliant XML schema is summarized in https://www.seegrid.csiro.au/twiki/bin/view/AppSchemas/HollowWorld and

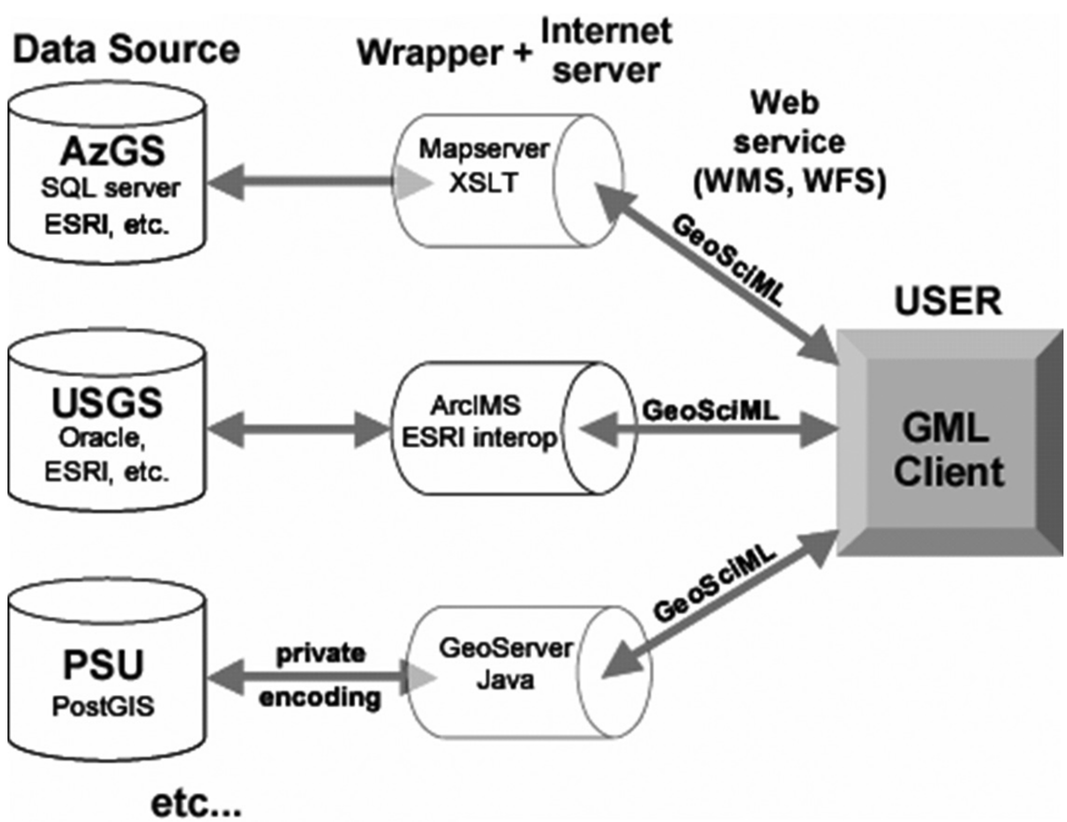


Figure 2. Communication between data providers and consumers utilizes standard GeoSciML schema. Clients that can interpret GeoSciML can operate with any GeoSciML-enabled data source.

http://www.seegrid.csiro.au/twiki/bin/view/AppSchemas/OandMCookbook. See also Boisvert et al. (2004) from the USGS DMT 2004 workshop. Use of a standard graphical notation for model representation during development makes group analysis and review of the evolving model much easier.

Major Entities

Only a small part of the GeoSciML model is discussed here. See the GeoSciML Twiki (https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/GeoSciML) for more information about the full model. Figure 3 presents the logical framework that underlies the draft GeoSciML version 2 GeologicFeature implementation, a core aspect of GeoSciML. Starting from the center left, a MappedFeature associates a GeologicFeature with a GML_geometry that specifies a location on or within the Earth. The mapped feature may be the result of an Observation if observation-related metadata concerning identification of the mapped feature are recorded. Each GeologicFeature is associated with a ControlledConcept classifier that specifies the intention of what the GeologicFeature represents. A GeologicFeature may have one or more associated GeologicFeatureDescriptions that specify properties assigned to the feature. Each description may also be represented as the result of an Observation. Table 1 summarizes the packages included in the GeoSciML UML model. Each package is implemented as a separate XML schema.

Geologic Feature

In the draft GeoSciML version 2 model, GeologicFeature is an association class that binds mapped feature(s) and description(s) with one or more classification concepts. Geologic feature is an entity that represents some particular phenomenon that may be observed in the Earth. It has a primary classification in terms of a controlled concept, and this association establishes a content model or concept space within which the feature is located/given identity by specification of a collection of properties in a description. A MappedFeature instance specifies a particular located occurrence of a geologic feature by associating it with a location (GML_geometry). GeologicFeatures may be classified by geologic unit or geologic structure ControlledConcepts terms. In addition to its primary classification (e.g. a lithostratigraphic designation), a feature may carry alternative classifications (e.g. geotechnical classification). GeologicFeature corresponds with a "legend item" from a traditional geologic map and with "occurrence" in conceptual models presented by Brodaric and Gahegan (2006) or Richard (2006). GeologicFeatures may have one or more associated GeologicFeatureDescriptions. Multiple descriptions associated with a feature may be the result of different observations (different observer, different time, different observation procedure…), or may specify different properties

Mapped Feature

A MappedFeature is a specific bounded occurrence, such as an outcrop or map polygon that carries a geometry or shape (through its samplingFrame association). It has an associated GeologicFeature instance that specifies what kind of thing is represented by the mapped geometry, both by classification with a vocabulary term (ControlledConcept) and through association with one or more description objects (GeologicUnitDescription) that specify property values.

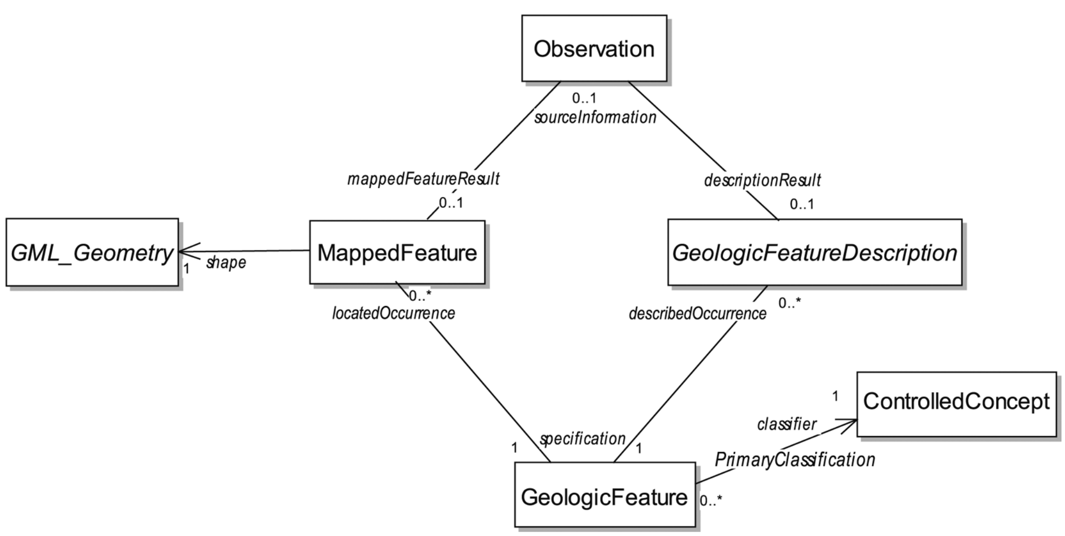


Figure 3. Core GeoSciML 2.0 logical model.

Table 1. Packages in GeoSciML model.

Package Name	Contents
TopLevel	The core model for mapped entities distinguishes between geologic features, mapped features, and controlled concepts
BasicTypes	Simple package, extends gml:MeasureType to represent quantification of measurements using relative comparisons, e.g. greater than, less than.
LiteralValue	The GeoSciML "value" model provides a generic way of encoding "literal" values, both textual and numeric, which have uncertainty and may be a range. These values are usually obtained as the result of an observation. The description of the associated observation event will provide more detail about the observation method, result quality, etc.
RootDoc	Generic collection element for packaging objects from the GeoSciML schema.
GeologicMetadata	Interim model for representation of dataset, feature, and attribute-level metadata. ISO 19115 metadata would be preferred, but the XML implementation (ISO 19139) is not yet supported by common software The interim model is intended to have some forward compatibility with ISO. For example, the scope-codes are a subset of MD_ScopeCode from ISO 19115.
GeologicVocabulary	Model for controlled vocabularies of terms linked to normative descriptions, link to ontology. A GeologicVocabulary is a collection of terms (ControlledConcept) and relationships (VocabRelation).
BoreHole	Support for borehole data in GeoSciML is provided by XMML components. Borehole is modeled as a kind of sampling profile that may have various sorts of associated 'logs', modeled as kinds of coverages.
EarthMaterial	Earth Material is a class that holds a material description. A naturally occurring substance in the Earth. Earth Material represents substance, and is thus independent of quantity or location. Ideally, Earth Materials are defined strictly based on physical properties, but because of standard geological usage, genetic interpretations enter into the description as well.
GeologicAge	The age of a particular geological event or feature expressed in terms of years before present (absolute age), referred to the geological time scale, or by comparison with other geological events or features (relative age). A GeologicAge can represent an instant in time, an interval of time, or any combination of multiple instants or intervals. Specifications of age in years before present are based on determination of time durations based on interpretation of isotopic analyses of EarthMaterial (some other methods are used for geologically young materials). Ages referred to geological time scales are essentially based on correlation of a geological unit with a standard chronostratigraphic unit that serves as a reference. Relative ages are based on relationships between geological units such as superposition, intruded by, cross-cuts, or 'contains inclusions of'.
GeologicRelation	Geologic Relations are typed, directed associations between geologic objects. Represents any of a wide variety of relationships that can exist between two or more GeologicFeatures. 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. Two or more GeologicFeatures are associated in a GeologicRelation; each has a role in the relationship. Examples of geological roles include "overlies", "is overlain by", "is younger", "is older", "intrudes", "is intruded by", and so forth. In a relationship where an igneous unit intrudes a sedimentary unit, the geological relationship is 'intrudes', the intruded sedimentary unit has the role 'host', and the igneous unit has the role 'intrusion'. Many other types of relationships can also be accommodated via GeologicRelation, for example, topological relations between spatial objects could be described where they are scientifically significant.
GeologicTime	The GeoSciML Geologic Timescale model and encoding is described in detail in the paper 'A formal model for the geologic time scale and global stratotype section and point, compatible with geospatial information transfer standards' (Cox and Richard, 2005). The classic "geological time scale" is a hierarchical ordinal system, in which the eras are ranked: "stages" nest within "series" within "systems" within "eras" within "eons" (in the most common version of the ranking system).
GeologicUnit	Package containing content model for geologic unit. Geologic unit is a notional unit, whose complete and precise extent is inferred to exist. Practically, spatial properties are only available through association with a MappedFeature. Includes both formal units (i.e. formally adopted and named in the official lexicon) and informal units (i.e. named but not promoted to the lexicon) and unnamed units (i.e. recognizable and described and delineable in the field but not otherwise formalized).
StructureObject	Package containing content model for geologic structure. Version 1 includes fault system, fault, contact, and fault displacement.

Controlled Concept

ControlledConcepts represent human concepts in computer form, typically as words (lexical objects) with an associated definition. Because GeoSciML extends GML, each controlled concept instance may have one or more gml:names, but the GeoSciML model adds a preferred name element that specifies one term that is used to identify the concept. Practically speaking, each preferredName should be associated with a unique concept, but in a distributed system, this cardinality cannot be enforced. ControlledConcepts are aggregated into GeologicVocabulary collections, which are derived from a GML dictionary. Data producers should ensure that preferredNames are unique within a particular vocabulary. A Controlled Concept may have an associated prototype entity (not shown in Figure 3) that can be a GeologicFeature, EarthMaterial, or Specimen. The prototype entity provides a mechanism to associate machine-analyzable properties with ControlledConcept terms. Similar functionality might be provided by links from the ControlledConcept to some other formal ontology representation.

Geologic Feature Description

Descriptions are collections of properties with assigned values (e.g. attributes) that characterize some feature. Different kinds of descriptions specify different properties. Descriptions may be associated with Observation elements that supply information on the origin of the property value assignments.

Observation

Observation describes the "metadata" associated with an information capture event, together with a value for the result of the observation. Observations are the basis for classified features, interpretations, and models. GeoSciML uses the Observation and Measurement model from the Open Geospatial Consortium (Cox, 2006), which models observation as a kind of event, in which a result value is assigned to some property of a feature of interest, using some procedure.

SOME SCHEMA DESIGN AND USAGE PATTERNS ISSUES

Names and Identifiers

Any GML Object or Feature may have an unlimited number of gml:name properties, which reflects the fact that the same object often has different identifiers assigned by different authorities. To assert "this is the name or identifier assigned by authority XYZ corporation," use the codeSpace attribute on gml:name (i.e. the scope identifier). If the codespace for the gml:name is not specified, then the value is implicitly under the authority of the organization or service that supplies the document, which should be indicated by associated document-level metadata.

Note that GML document elements also include a gml:id attribute, which plays a different role from the gml:name element. The value of the gml:id has type="xsd:ID", so it must be unique within the (XML) document. It is a document fragment identifier that acts as a handle for an XML element in the scope of its appearance within a particular document, and is usually assigned by the information management system since it is primarily significant in that context. The gml:id supports cross-references within a document and references that involve individual nodes (elements) within a system of documents. The value of a gml:name has type="gml:CodeType", which is a string with a "codeSpace" attribute. In the context of a GML object, the value of a gml:name is a label or identifier for the object described by the containing element, and is typically assigned by the data provider agency. The gml:name should be used for identifiers that are required to be persistent and are subject to constraints (e.g. uniqueness) applicable to a context wider than just the document scope. Different authorities may have different authoritative identifiers for the same item.

Namespace and Packaging

The namespace for GeoSciML version 2.0 schema is http://www.cgi-iugs.org/xml/GeoSciML/2. Versioning strategy for namespace evolution will follow practice described in OGC 05-062r3. For future upgrades, each minor version of any such schema that retains the namespace of the predecessor shall not introduce any new XML types or elements that could not be safely ignored by existing application based on the previous minor version, which ensures a strong form of backward compatibility. Components from other namespaces (e.g., http://www.opengis.net/om) may also constitute a "canonical" part of GeoSciML but will be incorporated using the WXS import mechanism and, thus, retain their own namespace names.

The physical document location (path) for GeoSciML schema will include the complete version number—initially 1.0.0, moving to 1.0.x for bug-fix releases, and 1.1.x (etc.) for extensions that do not change the scope of the schema. Schema documents are hosted in the GeoSciML publish/build repository, which is at https://www.seegrid.csiro.au/subversion/xmml/GeoSciML/tags/.

Use of Scoped Names

Use of scoped names, i.e., a term or word with an identifier for the source of the term, provides a method for linkage to formal controlled vocabularies (e.g. an ontology) that may then be used for semantic mediation. For example, a GeoSciML file might have a property value specified by the following element:

```
<CGI_TermValue>
  <qualifier>common</qualifier>
  <value codeSpace=”http://www.iugs-
    cgi.org/outcropCharacterVocabulary”>
    ledge forming</value>
</CGI_TermValue>
```

The <value> element contains a scoped name “ledge forming” from the vocabulary specified by the codeSpace attribute. If the data interpreter is familiar with the “http://www.iugs-cgi.org/ outcropCharacterVocabulary” codeSpace (vocabulary), then they may use the scoped name directly or by correlation with a preferred term in a different vocabulary. On the other hand, if the identified codeSpace (vocabulary) is not familiar and its identifier is a resolvable URL that points to service that can provide a definition of the term in a known format (e.g. free text, OWL, KIF...), it is possible to interpret the term. This may be as simple as someone studying a free text definition and determining the closest corresponding term in their vocabulary. An automated semantic mediator might be able to use a formal definition (e.g., OWL) to match with the closest subsuming term in a different formal vocabulary that is preferred by the data interpreter.

Value specification

The GeoSciML data model includes a flexible value specification scheme that is designed to capture value descriptions conventionally recorded by geologists. All values may carry a qualifier. Numeric values include units of measure. Values may be specified in several manners:

- by a single numeric value with optional uncertainty, e.g., 5.24 +/- 0.12
- by a numeric range, e.g., 5.7-13.6
- by a term with an identifier for the source vocabulary, e.g., “thick-bedded (NADM SLTTs)”
- as a range with bounds assigned by terms or by a term and a numeric value, e.g. “fine- to medium-grained (Folk 1968)” or “Miocene (IUGS 2004)” – 1.7 Ma.

Instance Document Example

Example instance documents associated with each version of the schema in the subversion repository (https://www.seegrid.csiro.au/subversion/xmml/GeoSciML/tags/) are stored in an “Instances” subdirectory in the directory for that version. The following listing provides an example usage of many of the elements for geologic unit description. The base element in the document is a GeoSciML collection (gsml:); each member of the collection starts with a <member> element. GeoSciML collection members may be:

1. Geologic features (a kind of GML feature)
2. GML geometry elements
3. Mapped Features (outcrops, sample locations, traverses/sections)
4. Controlled concepts (vocabulary definitions)
5. Geologic relationships
6. Dictionaries (collections of controlled concepts)

Comments in the following listing are delimited by ‘`<!--`’ and ‘`-->`’.

```
<?xml version=”1.0” encoding=”UTF-8”?>
<Gsml xmlns=”http://www.cgi-iugs.org/xml/GeoSciML/1” ... other namespace declara-
tions>
   <!-- The lexicon would probably be in a separate file. The StratigraphicLexi-
con element extends GML dictionary (through GeologicVocabulary GeoSciML element)
-->
<member>
   <StratigraphicLexicon gml:id=”AZGSGeologicUnits”>
   <!-- This is a lexicon element that includes three units -->
      <gml:description>Collection of geologic units defined by State of Arizona</
                gml:description>
      <gml:name>Arizona stratigraphic unit lexicon</gml:name>
      <member>
          <ControlledConcept gml:id=”MartinFormationConcept”>
              <gml:description>lithostratigraphic formation defined by ... </gml:
                      description>
              <gml:name>urn:x-cgi:def:lithostratigraphy:USGS:2006:Geolex:Martin-
Formation </gml:name>
              <preferredName>Martin Formation</preferredName>
              <prototype xlink:href=”#Feature2524”/>
```

```
              <vocabulary xlink:href=”#AZGSGeologicUnits”/>
              <metadata/>
         </ControlledConcept>
      </member>
      <member>
         <ControlledConcept gml:id=”LS2”> ... </ControlledConcept>
      </member>
      <member>
         <ControlledConcept gml:id=”LS3”>   ...</ControlledConcept>
      </member>
   </StratigraphicLexicon>
</member>

<member>
   <GeologicFeatureRelation gml:id=”rel-100”>
   <!-- This is a geologic relationship element-->
       <gml:name>urn:x-cgi:def:lithostratigraphy:USGS:2006:featureRelation:
          Stratigraphic position</gml:name>
       <role codeSpace=”http://www.iugs-cgi.org/featureRelationVocabulary”>overli
          es</role>
       <source xlink:href=”#BeckersButteMemberPrototype”/>
       <target xlink:href=”#JeromeMemberPrototype”/>
   </GeologicFeatureRelation>
</member>

<member>
<!-- GeologicFeature is derived from GML AbstractFeature, it associates a de-
scription, a classifier (what is described) and an extent (where it was de-
scribed, if defined).The Classifier element defines the type of a feature. Mul-
tiple descriptions may be associated with a GeologicFeature -->
 <GeologicFeature gml:id=”Feature2524”> <!-- This is a geologic unit GML feature,
      which is the basic container for geologic unit descriptions in GeoSciML v.
      1 -->
   <gml:description>The type section of the Martin Formation at Mt. Martin near
            Bisbee consists almost entirely of  medium-gray to medium dark-gray
            aphantiic to fine-graine limestone. dolostone is entirely subordi-
            nate, ...
   </gml:description>
   <gml:name>urn:x-cgi:def:lithostratigraphy:USGS:2006:Geolex:TypeMartinForma-
tion</gml:name>
   <gml:boundedBy>
      <gml:Envelope>
        <gml:lowerCorner/> <!-- corners of a bounding box for type area of the
              Martin Formation; geometry specification elements not included
              here-->
        <gml:upperCorner/>
      </gml:Envelope>
   </gml:boundedBy>
   <purpose>typicalNorm</purpose>
   <age>
<!-- Geologic age element includes a date value specification (see below), and an
event specification that explicitly identifies the event to which the age is as-
signed (e.g. deposition, cooling through biotite closure temperature...) -->
      <GeologicAge>
        <value>
         <CGI_TermValue>
          <value codeSpace=”http://www.iugs-cgi.org/geologicAgeVocabulary”>Middle
```

```
                Devonian</value>
          </CGI_TermValue>
        </value>
        <event>
          <CGI_TermValue>
          <value codeSpace="http://www.iugs-cgi.org/EventVocabulary">deposition</
                value>
          </CGI_TermValue>
        </event>
      </GeologicAge>
   </age>
   <classifier xlink:href="#MartinFormationConcept"/> <!-- here's the link to the
          controlled concept that defines the intention of the Martin Formation.
          Link is reference to controlled concept instance in this document-->
   <description>
   <LithostratigraphicUnitDescription>
   <metadata/> <!-- xlink to metadata for this description; this provides tie to
         Observation model-->


   <partOf>
     <GeologicUnitDescriptionPart> <!-- 310- 340 thin bedded, non fossiliferous
            dolostone -->
      <unit>
        <LithostratigraphicUnitDescription gml:id="GeoUnitPart0235">
<!-- part is also a lithostratigraphic unit, uses same description schema as con-
taining unit; it could have parts itself; partonomy is recursive. -->
         <descriptionSource xlink:href="reference to description source observa-
                tion" /> <!—Source observation element not included here -->
         <bodyMorphology xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
         <outcropCharacter xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
         <grossGenesisTerm xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
         <exposureColor xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
         <grossChemistry>
            <CGI_TermValue>
               <qualifier>always</qualifier>
               <value codeSpace="http://www.iugs-cgi.org/grossChemistryList">carb
                         onate</value>
            </CGI_TermValue>
         </grossChemistry>
         <rank codeSpace="http://www.iugs-cgi.org/Vocabulary">DescriptionPart</
                rank>
         <weatheringCharacter xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
         <metamorphicGrade/>  <!-- not specified so implies same as containing
                unit -->
         <unitThickness>
            <CGI_NumericValue>
               <principalValue uom="meter">30</principalValue>
               <plusDelta uom="meter">20</plusDelta>
               <minusDelta uom="meter">10</minusDelta>
            </CGI_NumericValue>
         </unitThickness>
         <beddingStyle xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
         <beddingPattern xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
         <beddingThickness>
           <CGI_TermValue>
             <value codeSpace="http://www.iugs-cgi.org/ThicknessVocabulary">Thin-
```

```
                         bedded</value>
            </CGI_TermValue>
          </beddingThickness>
         </LithostratigraphicUnitDescription>
      </unit>
      <role codeSpace=”http://www.iugs-cgi.org/unitPartRoleVocabulary”>Stratigrap
               hic part</role>
      <type>codeSpace=”http://www.iugs-cgi.org/unitPartTypeVocabulary”>Descriptiv
               ePart</role>
      <proportion>
         <CGI_NumericValue>
         <qualifier>approximate</qualifier>
         <principalValue uom=”percent”>12</principalValue>
         <plusDelta uom=”percent”>0</plusDelta>
         <minusDelta uom=”percent”>0</minusDelta>
         </CGI_NumericValue>
      </proportion>
     </GeologicUnitDescriptionPart>
   </partOf>
<!-- end of part descriptions. Following properties apply to entire described
unit -->

   <descriptionSource xlink:href=”reference to description source observation” />
   <bodyMorphology xlink:href=”urn:x-ogc:def:nil:OGC:unknown”/>
   <outcropCharacter>
      <CGI_TermValue>
         <qualifier>common</qualifier>
         <value codeSpace=”http://www.iugs-cgi.org/outcropCharacterVocabulary”>le
              dge forming</value>
      </CGI_TermValue>
   </outcropCharacter>
   <grossGenesisTerm
      <CGI_TermValue>
         <qualifier>always</qualifier>
         <value codeSpace=”http://www.iugs-cgi.org/GenesisVocabulary”>Sedimentary
               , marine</value>
      </CGI_TermValue>
   </grossGenesisTerm>
   <exposureColor>
      <CGI_TermValue>
         <qualifier>common</qualifier>
         <value codeSpace=”http://www.color.org/ColorVocabulary”>Light gray</val-
              ue>
      </CGI_TermValue>
      <CGI_TermValue>
         <qualifier>common</qualifier>
         <value codeSpace=”http://www.color.org/ColorVocabulary”>Medium gray</
            value>
      </CGI_TermValue>
      <CGI_TermValue>
         <qualifier>rare</qualifier>
         <value codeSpace=”http://www.color.org/ColorVocabulary”>Pink</value>
      </CGI_TermValue>
   <exposureColor/>
   <grossChemistry>
      <CGI_TermValue>
```

```
        <qualifier>common</qualifier>
        <value codeSpace="http://www.iugs-cgi.org/grossChemistryList">carbonate</
                value>
      </CGI_TermValue>
      <CGI_TermValue>
        <qualifier>occasional</qualifier>
        <value codeSpace="http://www.iugs-cgi.org/grossChemistryList">siliceous</
                value>
      </CGI_TermValue>
   </grossChemistry>
   <rank codeSpace="http://www.iugs-cgi.org/Vocabulary">Formation</rank>
   <weatheringCharacter xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
   <metamorphicGrade>
     <CGI_TermValue>
       <qualifier>always</qualifier>
       <value codeSpace="http://www.iugs-cgi.org/metamorphicGradeVocabulary">not
            metamorphosed</value>
     </CGI_TermValue>
   </metamorphicGrade>
   <unitThickness>
      <CGI_NumericValue>
         <principalValue uom="meter">340</principalValue>
         <plusDelta uom="meter">10</plusDelta>
         <minusDelta uom="meter">10</minusDelta>
      </CGI_NumericValue>
   </unitThickness>
   <beddingStyle>
      <CGI_TermValue>
         <qualifier>common</qualifier>
         <value codeSpace="http://www.iugs-cgi.org/Vocabulary">Planar bedding</
                     value>
      </CGI_TermValue>
   </beddingStyle>
   <beddingPattern xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
   <beddingThickness>
      <CGI_TermRange>
        <lower>
         <CGI_TermValue>
            <qualifier>common</qualifier>
            <value codeSpace="http://www.iugs-cgi.org/Vocabulary">thin bedded</
                           value>
         </CGI_TermValue>
        </lower>
        <upper>
         <CGI_TermValue>
            <qualifier>rare</qualifier>
            <value codeSpace="http://www.iugs-cgi.org/Vocabulary">thick bedded</
                           value>
         </CGI_TermValue>
        </upper>
      </CGI_TermRange>
   </beddingThickness>
   </LithostratigraphicUnitDescription>
   </description>
    </GeologicFeature>
  </member>
</Gsml>
```

TEST BED DEMONSTRATION

Six national and two state geological survey agencies, in Australia, Europe, and North America, participated in a proof-of-concept demonstration of GeoSciML at the International Association of Mathematical Geologists (IAMG) meeting in Liege, Belgium, in September 2006. The demonstration showed that it is possible to access information in real time from globally distributed data sources. Geological map polygons and attribute information, and borehole data, were displayed, queried, and re-portrayed using web applications hosted by the Geological Survey of Canada and the French Bureau de Recherche Géologiques et Minières (BRGM). Functions demonstrated included continuous map portrayal with attribute query, reclassification according to attributes, and download of complex data structures encoded in GeoSciML.

Information delivery from different complex data stores using a community standard schema demonstrated that GeoSciML provides a data model and format capable of supporting transfer of geology data from multiple jurisdictions. This also demonstrated that a distributed data delivery system can be constructed by specifying standard interfaces, not limited to single vendor software. New services can be added easily, providing they conform to the interface. All of the services in the test bed used different data stores, wrapped by a variety of server software applications. Deployment requires configuration of server- and client-side software to conform to the data model, but does not require development of new software "from scratch."

Three use cases were demonstrated at the IAMG 2006 meeting in Belgium. Use Case 1 demonstrated display of map data and query for the description of a single map object. When the client asks for the map, the server returns a map with default symbolization. A user can then click on any graphic feature from a layer to retrieve information for the feature, which can be presented to the user as raw GeoSciML or as a more clearly-rendered HTML version. Presentation formats other than HTML can be requested by the client if the server supports them. The types of features used must include at least one of the following: geologic units, faults, contacts or boreholes.

Use case 2 demonstrated selection and download of features; a geographic bounding box is specified and the contents downloaded as a GeoSciML document. The GeoSciML document can be reformatted (e.g. by XSLT for display in a browser) or serve as input for another process in a workflow. The GeoSciML document contains a collection of GeologicFeatures or Boreholes.

Use case 3 demonstrated dynamic query and re-symbolization of mapped features on the basis of age, using the IUGS standard geologic age color scheme, or on the basis of lithology, using a CGI defined lithology color scheme. The results of symbolization by lithology for data from Canada, the U.S., and Scandinavian countries is shown in Figure 4. A very simple lithologic classification and symbolization was used, with four classes and related colors: igneous (pink), sedimentary (green), metamorphic (purple), and unconsolidated (yellow). Each participant had to implement a mechanism to map from properties associated with the mapped features to the standardized lithology classes. It is the service provider's prerogative to determine the mapping from the data source to the classification.

SUMMARY

A standardized schema and syntax for information encoding is a fundamental requirement for interoperable infor-

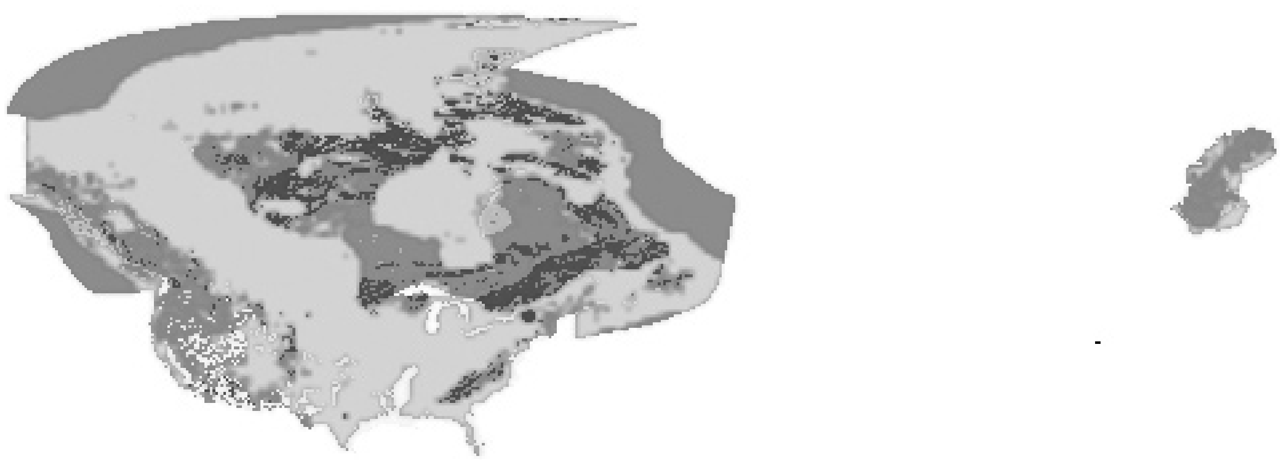


Figure 4. Use Case 3 from Testbed 2, re-symbolization of geologic units by lithology for Canada, U.S. and Scandinavian countries: igneous (medium gray), sedimentary (light grey), metamorphic (dark gray), and unconsolidated (nearly white).

mation systems. The IUGS CGI Data Model collaboration working group has developed GeoSciML, an XML-based GML (geography markup language) application, to meet this requirement for the interchange of geoscience information. The schema for this application reuses existing markup languages where possible. Newly developed markup specifications are based on existing conceptual models in most cases. This standards-based data format provides a framework for application-neutral encoding of geoscience thematic data and related spatial data. It is intended for use in publishing or interchanging data between organizations that use different database implementations and software/systems environments. Full realization of data interoperability at the semantic level will require development of controlled vocabulary resources for specifying actual content. A Testbed demonstrated simple interoperability using web map and feature services (WMS, WFS) between geological surveys in several different countries. GeoSciML is being considered as a national standard for geoscience data exchange by federal and state geological surveys in Australia and the European Union Spatial Data Infrastructure (INSPIRE), and will be submitted in 2007 as an IUGS-CGI specification.

Development of GeoSciML is an open process with the intent to involve as many participants as possible. This will ensure development of a schema and services that will meet the needs of a wide variety of geoscience data producers and users. Three types of participation are available: 1) direct participation in GeoSciML development, 2) monitoring GeoSciML development via the web-collaboration tools and 3) deploying an internet server to provide data in GeoSciML format.

REFERENCES

Boisvert, E., Johnson, B.R., Cox, S.J., and Brodaric, B.M., 2004, GML Encoding of NADM C1, in Soller, D.R., ed., Digital Mapping Techniques '04—Workshop Proceedings: U.S. Geological Survey Open-File Report 2004–1451, p. 95-103. Available at http://pubs.usgs.gov/of/2004/1451/boisvert/index.html (accessed 12/12/2006).

Brodaric, B.M., and Gahegan, Mark, 2006, Representing geoscientific knowledge in cyberinfrastructure: some challenges, approaches and implementations, in Sinha, Krishna, ed., Geoinformatics-Data to knowledge: Geological Society of America Special Paper 397, p. 1-20.

Cox, S.J.D., ed., 2006, Observations and measurements: Open Geospatial Consortium, Inc., document OGC 05-087r4, version 0.14.7, 168 pages. Available at http://portal.opengeospatial.org/files/?artifact_id=17038 (accessed 12/13/2006).

Cox, S.J.D., Daisey, P., Lake, R., Portele, Clemens, and Whiteside, Arliss, eds., 2004, OpenGIS Geography Markup Language (GML) Implementation specification: Open GIS Consortium, Inc, document OGC 03-105r1, version 3.1.0, 580 p.

Cox, S.J.D., and Richard, S.M., 2005, A formal model for the geologic time scale and global stratotype section and point, compatible with geospatial information transfer standards: Geosphere, v. 1, p. 119-137, DOI: 10.1130/GES00022.1. Available at http://geosphere.geoscienceworld.org/cgi/content/abstract/1/3/119 (accessed 12/12/2006).

North American Geologic-Map Data Model Steering Committee, 2004, NADM Conceptual Model 1.0, A Conceptual Model For Geologic Map Information: U.S. Geological Survey Open-File Report 2004-1334, 61p., available at http://pubs.usgs.gov/of/2004/1334/.

Richard, S.M., 2006, Geoscience concept models, in Sinha, Krishna, ed., Geoinformatics-Data to knowledge: Geological Society of America Special Paper 397, p. 81-108.

SC36 Secretariat, 2003, Proposed Draft Technical Report for: ISO/IEC 2382, Information technology—Learning, education, and training—Management and delivery—Specification and use of extensions and profiles: ISO/IEC 2382-01, ISO/IEC JTC1 SC36 N0646, available at http://jtc1sc36.org/doc/36N0646.pdf (accessed 12/12/2206).

Open Source Web Mapping: The Oregon Experience

By David Percy

Portland State University
Department of Geology
1721 SW Broadway
Portland, OR 97201
Telephone: (503) 725-3373
Fax: (503) 725-3325
e-mail: percyd@pdx.edu

SUMMARY

Open source tools have enabled a superior level of productivity for our online interactive map data delivery efforts. At Portland State University (PSU), we have delivered a web-accessible interactive map of the geology of Oregon since 1999. We also began delivering glacier data and coastal data in the ensuing years as it became obvious that all scientific data needed a web presence. Initial versions of our interactive web maps were based on MapObjects, followed by ArcIMS, both of which are products of ESRI, the dominant provider of proprietary GIS software. As more scientific data needed to be delivered via the Internet, the limitations of ESRI's software became evident, and the search began for better solutions.

Eventually, a new direction was established using Open Source software, which allows for a greater degree of customization and the transfer of existing skills in web design. Since PSU has an expert group of developers that use open source tools, such as PHP and MySQL, the training was minimized. Many of the skills that are already used for other web applications are directly transferable to web mapping when we use open source tools. Additionally, we have a culture of Open Source software use and development at PSU and, in general, within the state of Oregon, so this direction of development makes sense on many levels.

Overall, this effort has resulted in a web-mapping framework that provides considerably faster web page updates than its proprietary counterparts. Additionally, we have the ability to re-use components from applications, developed initially for certain organizations, to solve new problems for other organizations. Each time we initiate a project, we consider how it can benefit the larger goals of the web-mapping framework, which is essentially to provide feature for feature replacement of our proprietary competition, and thus everyone benefits.

BACKGROUND

Since Linux and Apache are supported on our campus as the defacto web platform, it is obvious that, to minimize the support required, we should use this platform. Mapserver was developed by researchers at the University of Minnesota under an NASA grant and is a mature server side application for delivering map data. PostGreSQL is an open source project that traces its roots to UC Berkeley, but is currently maintained in Germany. It is a hybrid relational-object oriented database, similar in functionality to Oracle. PostGIS, which was developed by Refractions Research (an open source consulting company in Victoria, BC), is a set of extensions that enhance PostGreSQL to give it a full set of GIS capabilities. PostGIS implements the full set of "OpenGIS Simple Features for SQL" capabilities as specified by the Open Geospatial Consortium. We refer to this mapping platform as LAMP for Linux/Apache/Mapserver/PostGIS.

METHODS

Once an organization has decided to use this set of open source mapping tools, they must decide how to deliver the data in an interactive web application. The previously enumerated tools provide the back end for web data delivery, but a front-end is needed to allow the end-user to interact with the data in a web browser.

Several mature web-mapping frameworks exist, but on close examination, it was clear that some had the patina of an older web application. That is, applications on the internet mature and age quickly, and new developments also happen quickly—the term "internet time" has currency because it is true that things happen rapidly on the internet. Thus, even though we could use one of the existing map-frameworks, that would not mean that it would be as functional as something developed with an eye to the future.

Initially, we considered using one of the existing web-mapping frameworks. There are several robust applications in existence such as Chameleon, Mapbuilder, and Ka-Map. After examining these mapping frameworks, which were already deployed, and in light of the previous notes and the ascendency of Web 2.0 and AJAX, we decided to develop our own. The functional requirements were to zoom/pan, query, turn off and on thematic layers, and dynamically resize the window to maximize map area. In a single weekend of development, several PSU graduate students wrote a new framework. One year later, the resulting product was named Map-Fu and became its own open-source project on Sourceforge in December, 2006 (http://sourceforge.net/projects/map-fu/).

In the meantime, several other mapping front-end products have become available that provide the same types of features we developed in Map-Fu. Thus, there are many options for the open source enthusiast to pursue. The main cautionary note is for the potential user to follow the listserve of any particular project for a few weeks to determine how active the community is. A healthy open source project will have several posts to the listserve every day, usually even 10 to 20. A project that has not had any posts to the listserve for more than a month is probably dead or perhaps mature, yet used only by one group.

Regardless of what front-end an organization chooses, the first step is to develop a mapfile Mapserver can read and generate images from. The mapfile will consist of names of data sources such as shapefiles for vector data and geotiffs for basemaps. It will also specify how to symbolize individual classes of data, for example a "Qal" unit would likely be displayed with an RGB value of 255 255 0.

In terms of optimizing data for web delivery, a few tasks are required. Large raster data sets need to be tiled and have internal overviews built. This is done via a series of command line operations that utilize an open source library known as GDAL. To build internal tiles we issue a command like this:

```
gdal_translate -of GTiff -co "TILED=YES" shaded_relief.tif shaded_relief_tiled.tif
```

After this, it is useful to build internal overviews (similar to "pyramids" in ArcMap) using a command like:

```
gdaladdo shaded_relief_tiled.tif 2 8 32 128
```

Note that the first command, gdal_translate, creates a new file, while the second command, gdaladdo, works "in situ" (without creating a new file). Also, the execution order of commands matters. Overviews are not copied during a gdal_translate operation, so the user should build tiles first, followed by overviews, as illustrated above.

To optimize the vector data, the shapefiles are imported to the open source database PostGIS, which is an extension of PostGreSQL. From here a command line function is executed that produces a lower resolution data set for initial delivery at low resolution ("zoomed-out") levels. It uses the conversion from postgresql to shapefile with the addition of an SQL operation. In this example, we have a table named lithology in the database named geology. We request that the output be a shapefile name simplelith and the sql command simplify the vertices down to one every 1000 feet:

```
pgsql2shp -f simplelith -h localhost -u mapserve geology -s "select simplify(the_geom, 1000) as the_geom, gnlith_u from lithology"
```

Techniques like this can considerably speed up the delivery of web-accessible data. While it may seem strange to do such things, it is simply the reality of providing data on the internet, where delivery times mean the difference between users accessing your site or simply abandoning it for lack of responsiveness.

With regard to returning query results, we have implemented an approach that uses the geospatial database PostGIS. When the user clicks on the map with the "info-query" tool, the coordinate pair they clicked on is sent to the database, which then returns all objects from all tables that intersect the point that was clicked. We then have a query handler that outputs data related to the objects, depending on which layers in the view are on or off.

It is relatively trivial at this point to set up an Open Standards based output system. Note that Open Standards are different from Open Source; in the first case we are talking about a committee of vendors and organizations that decide upon a protocol for data interoperability, while in the latter we refer to a formal system by which users are allowed to view and legally modify and redistribute the source code of programs.

The Open Standards protocols that are of interest are Web Map Services (WMS) and Web Feature Services (WFS), though the suite of open standards-based web services are collectively called Open Web Services (OWS). The entire set of OWS is still under development, though it is maturing rapidly and there are some viable uses now. By inserting certain metadata statements into the same mapfiles we use for our interactive maps, we can simultaneously serve as OWS providers. This allows our data to be aggregated by others into other useful web-interfaces.

CONCLUSION

We have had great success and satisfaction using open source tools for our web delivery of scientific data. Using open source tools has given us the ability to leverage existing strengths, as opposed to having to learn techniques that only apply to one monolithic proprietary software program. We also have the ability to use data from other OWS providers, such as NASA or the USGS, as base layers on which to overlay our data. In the end, we assume that these “stove pipe” solutions we are building will be converted into pure open standards based formats like WMS and WFS, such that any standards-based interface can integrate our data with whatever other data they deem useful to addressing the situation at hand.

Lidar Basics for Mapping Applications

By James D. Giglierano

Iowa Geological Survey
Iowa Department of Natural Resources
109 Trowbridge Hall
Iowa City, IA 52242-1319
Telephone: (319) 335-1594
Fax: (319) 335-2754
e-mail: jgiglierano@igsb.uiowa.edu

INTRODUCTION

The purpose of this article is to help newcomers understand the basics of lidar data collection and processing, especially non-engineering, mapping specialists such as geologists, soils scientists, and those interested in land cover characterization. Many states in the U.S. are embarking on large-scale lidar acquisitions. This will make lidar elevation and other derived products widely available to many different audiences. To make full use of this new source of information, mappers must have or acquire some knowledge of the lidar data collection and handling procedures, and have the capability to convert the vendor supplied files into useful products. In some cases, mappers will do the processing themselves; in others, they will opt to have the processing performed by a vendor or third party. Another case may be that lidar derived topographic data supplied by a local government entity will have no metadata. In this case, the user will have to make some educated guesses as to the type of processing that may have been performed on the data.

HOW LIDAR DATA ARE COLLECTED AND REPRESENTED

The term **LiDAR** is an acronym for Light Detection and Ranging. Light Detection and Ranging basically consists of a laser rangefinder that operates in some form of airborne platform (helicopter, plane, or satellite). The rangefinder takes repeated measurements of the distance from the platform to the ground. The position and elevation of the platform is precisely known by way of airborne GPS along with ground control, so the elevation of the ground surface can be calculated by subtracting the laser rangefinder distance from the height of the platform. Compensation must be made for the tilt and pitch of the airborne platform by way of gyroscopes and accelerometers in the aircraft's inertial measurement unit. A good technical overview of lidar scanning technology is provided by Wehr and Lohr (1999).

Lidar systems record thousands of highly accurate distance measurements every second (newer systems operate at frequencies up to 150 kHz; older systems 30-80 kHz) and create a very dense coverage of elevations over a wide area in a short amount of time. Because lidar is an active sensor that supplies its own light source, it can be used at night and, thus, avoid routine air traffic, or it can be flown under some types of high cloud conditions. Most lidar systems record multiple surface reflections, or "returns," from a single laser pulse. When a laser pulse encounters vegetation, power lines, or buildings, multiple returns can be recorded. The first return will represent the elevation near the top of the object. The second and third returns may represent trunks and branches within a tree, or understory vegetation. Hopefully, the last return recorded by the sensor will be the remaining laser energy reflected off the ground surface, though at times, the tree will block all the energy from reaching the ground. These multiple returns can be used to determine the height of trees or power lines, or give indications of forest structure (crown height, understory density, etc.). Figure 1 shows a single 2 x 2 km tile that consists of 3.3 million first return lidar points.

Another feature of an airborne lidar system is the use of mirrors or other technology to point the laser beam to either side of the aircraft as it moves along its path. Depending on the scanning mechanism, the lidar scans can have a side-to-side, zigzag, sinusoidal, or wavy pattern. While the laser itself pulses many thousands of times per second, the scanning mechanism usually moves from side-to-side at around 20-40 cycles per second. This scanning, combined with the forward motion of the aircraft, produces millions of elevations in a short distance and time. The field of view or angle the scan makes from side-to-side can be adjusted by the operator, but is usually set at 30 to 40 degrees. This creates a swath of around 1 kilometer wide or less. Adjacent swaths overlap from 15 to 30% so that no data gaps are left between flight lines.

The spacing of lidar points on the ground, called "postings," is a function of the laser pulse frequency, scan

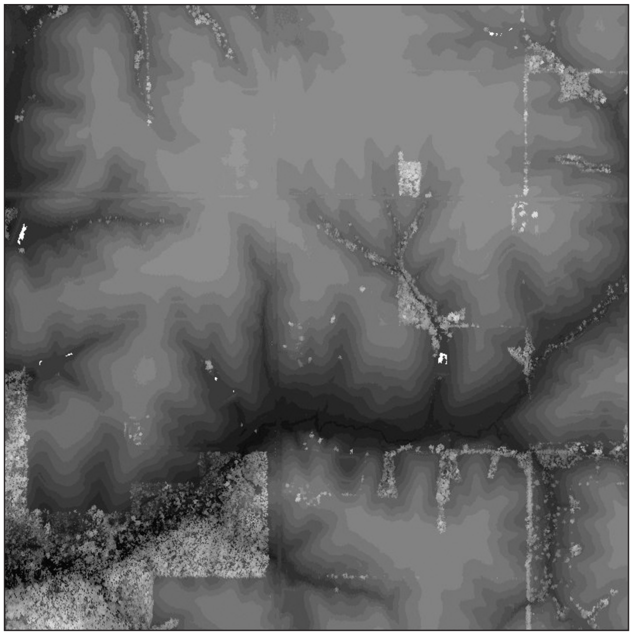


Figure 1. Gray scale image consisting of 3.3 million lidar first return points or “postings.” First returns indicate the tops of trees and buildings as well as bare ground in open areas. White areas are data voids where no returns were recorded, usually due to non-reflecting water surfaces.

frequency, and flight height (Baltsavias, 1999). While there is usually a nominal posting spacing specified in a lidar project, actual data points have variable spacing that are smaller and larger than the specified spacing. Mappers need to be aware of these effects when viewing final products that were derived from the raw data. The second aspect is that, because the laser scans from side to side, it interacts with the ground in different ways, depending on the angle of incidence. Lidar pulses at the edge of a scan will strike the sides of buildings, whereas pulses at the center of a scan will only strike the roof tops. Likewise, pulses at the edges of scans will pass through trees at an angle. Sometimes this will create “shadows” on the other side where no lidar passes through. In addition, less energy will return to the lidar receiver as it reflects away from the aircraft. This is evident in the intensity images created from the intensity values for each return: one can see overall darkening of the intensity at the edges of swaths. Edges of swaths appear darker than the returns at centers of swaths.

HOW LIDAR POINTS ARE PROCESSED INTO TINS AND DEMS

In the Spring of 2005, the Iowa Department of Natural Resources (DNR) and others, had lidar with a nominal resolution of 1 meter collected by a commercial vendor over the Lake Darling watershed located in Washington County, Iowa. The vendor’s lidar system collected a first and last return from each lidar pulse. From the first and last returns, a so-called “bare earth” return was created using a proprietary classification algorithm developed by the vendor. These classification systems try to sort out non-bare earth returns (tree tops, buildings, power lines, automobiles) from bare earth returns. To distinguish bare earth in forested areas, differences in elevation between the first and last returns, relative changes in elevation, and slope were used. Intensity data were used to identify vegetation and man-made materials. The lidar data for the Lake Darling watershed were collected in April with mostly leaf-off conditions. There are some data voids in forested areas due to non-penetration of the laser through tree canopy, but these areas are generally less than 10 meters across and are easily filled in by interpolation. Leaf-on conditions and tall crops, such as corn, do not allow easy penetration of the laser beam to the ground and should be avoided.

Lidar data for the Lake Darling area were supplied by the vendor in ASCII text format, consisting of 2 x 2 kilometer tiles with x and y coordinates, z elevations, and intensity values. With a nominal 1 meter posting spacing, some tiles had up to 3.3 million points. Postings near the center of the flight lines were close to the nominal 1 meter spacing (Figure 2), while toward the ends of scans, the points converge with the start of the next scan (Figure 3). While some scans converge, others diverge. Where the scans converge, the points can be less than half of the nominal spacing, and likewise, where they diverge, they can be twice the nominal spacing. Because some points can be as close as 0.5 meters, the tiles were initially interpolated to create grids with 0.5 meter resolution, with the idea that no data points should be merged or averaged with any other points. There is a tendency among some users to create grids with resolutions of 3, 5, and even 10 meters to save storage space or reduce the volume of data to process. We desired to create the grids as close as possible to the native resolution of the lidar data to evaluate their full potential to represent the smallest surface features.

To make digital elevation models (DEM) from the tiles, the Surfer 8 software (http://www.goldensoftware.com/products/surfer/surfer.shtml) was used. This software first creates a triangulated irregular network (TIN) before it interpolates the points into a raster DEM; however, once the DEM tiles were initially put together into mosaics, it became obvious that there were noticeable gaps between each tile. To remedy this, a C program was created to sort through the ASCII text files of the adjacent tiles and find points within a 3 meter buffer of the edge of the tile to be processed. Then the tiles were reprocessed adding the 3 meter buffers. When these raster tiles were merged together into a mosaic, the gaps were almost completely eliminated. Leica Imagine (http://gi.leica-geosystems.

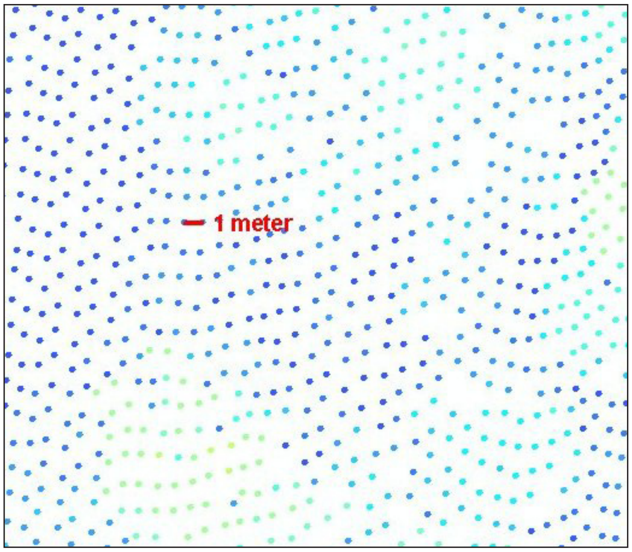


Figure 2. Data near the middle of a lidar flight line. Posting spacing is around one meter at the center of back and forth scans.

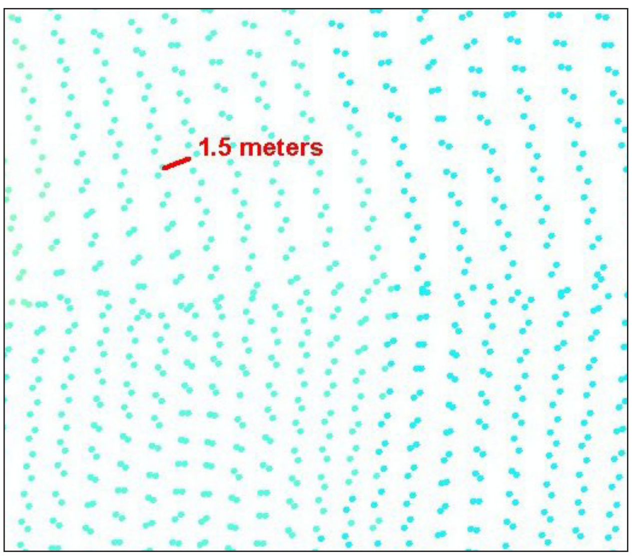


Figure 3. Data along overlap of two adjacent lidar flight lines (top and bottom of figure). The posting spacing is highly variable at edges of flight lines. Some postings are less than one meter apart at the end of one and beginning of the next scan, while the distances between points in different sets of scans are as much as 1.5 meters apart.

com/) was used to mosaic all the tiles into one large raster DEM file. From the DEM, shaded relief images were created and compressed. DEMs and shaded relief images were then easily imported into ArcGIS software (http://www.esri.com/) for display and further analysis.

Field examination of the lidar bare earth shaded relief images was conducted in January, 2006. It was surprising how well lidar shaded the relief images represented the smallest topographic features, including small slope changes of less than half-a-meter, even in forested areas. There were some data voids due to lack of penetration through the dense tree canopy, but there were enough data points to show good definition of incised stream channels, meander scars, and gullies (Figure 4). Man-made features such as road ditches and embankments, terraces, and dams were also well defined. Tillage patterns are evident as regular lineated textures on crop fields parallel to the road grid. These are not scanning artifacts as the individual scans are at a slight angle to the east-west flight lines.

Because the bare earth processing does not remove 100% of the forest artifacts, a distinctive bumpy pattern is left in the bare model, which indicates the presence of forest cover (Figure 5). During field examination, it was noticed that different canopy structures were represented by different patterns in the artifacts. In the tall canopy floodplain forest, most of the bumps were removed, which left a mostly smooth surface, while on side slopes with thick understory or brush cover, the texture on the shaded relief image is rougher in appearance. Interestingly, the bare earth processing removed nearly all of the numerous tree falls in the stream channels, which allows drainage tracing programs to work well when following flow paths downstream. Also, areas with pine trees were very smooth, which indicated nearly complete penetration by the laser beam.

HOW TO USE LIDAR PRODUCTS FOR MAPPING APPLICATIONS

Once the raw lidar point tiles are processed into high-resolution DEMs, other useful mapping products can be derived. The derived shaded relief image previously mentioned (Figure 4) is very useful for visual display and interpretation, and can be combined with colorized elevation images for extra information content. Another useful display product is the slope map, which can be derived from the DEM using the grid processing tools found in almost every GIS package. Usually, a choice can be made whether to calculate the slope rate in degrees or as percent (45 degree slope = 100%). A slope map based on percent can be grouped into slope classes typically used by soil survey mappers (slope class A = 0-2%, B = 2-4%, etc.) and readily compared to soil polygons displayed by slope class (Figure 6). Figure 7 shows the new level of detail available in slope classes derived from lidar data.

In addition to the elevation component of the lidar return, many systems produce an intensity component that indicates the strength of the lidar return. This intensity value is mostly influenced by the reflectance of the material struck by the laser pulse, but is also influenced by the scan angle. (Laser pulses directed away from the airplane at significant angles do not reflect back as much light energy as a pulse directed straight down from the plane). Because most lidar systems use a laser that emits light in

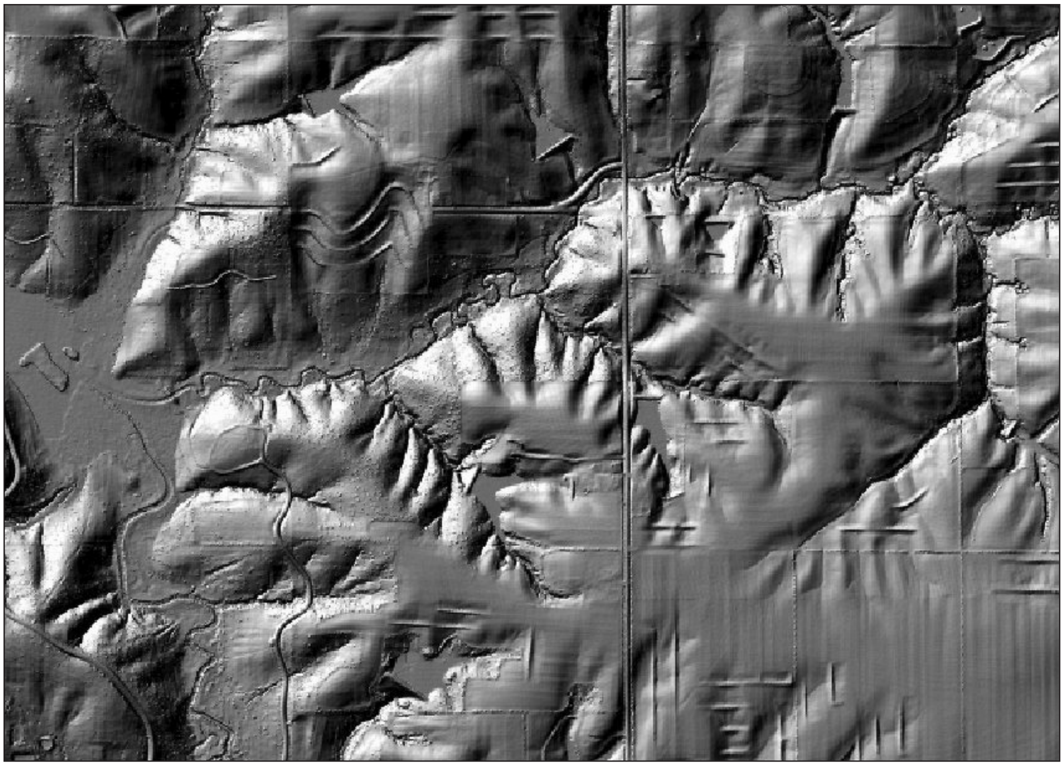


Figure 4. Portion of the bare earth shaded relief image of the Lake Darling watershed, showing natural and man-made features readily apparent in the lidar data.

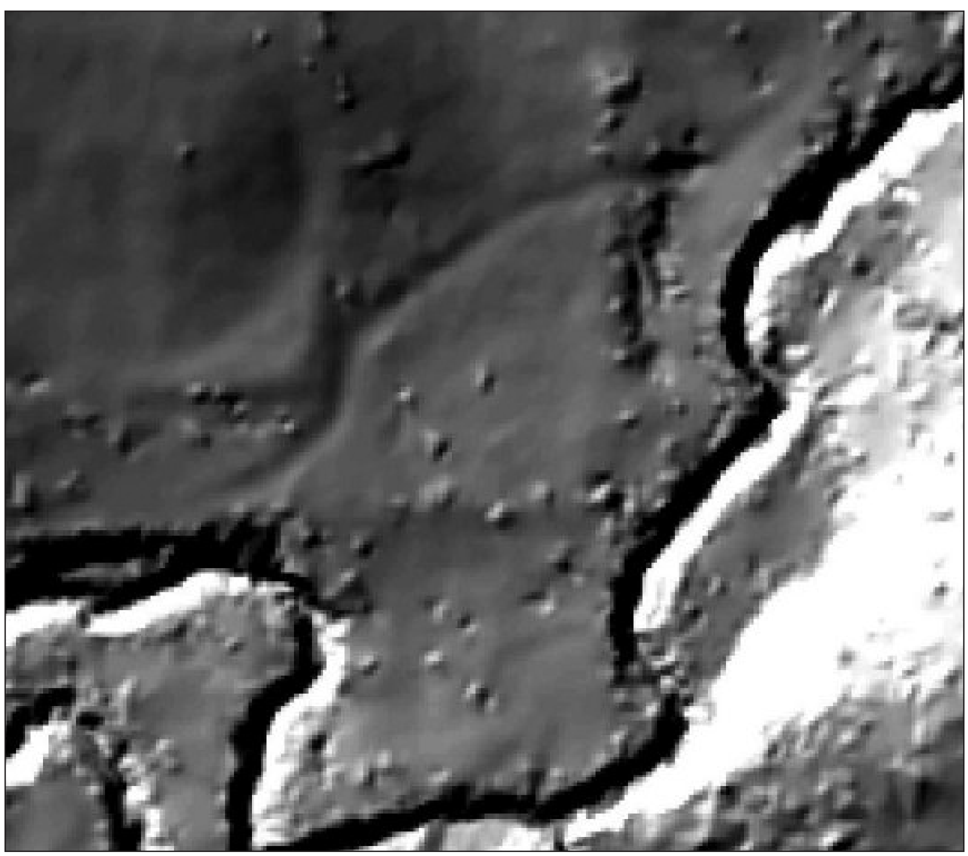


Figure 5. Portion of a bare earth shaded relief image showing artifacts (bumpy texture) in deciduous forest areas. These artifacts are lidar elevations classified as bare earth, but probably are from tree trunks, branches, or understory close to the ground and classified as bare earth by the vendor's algorithm.

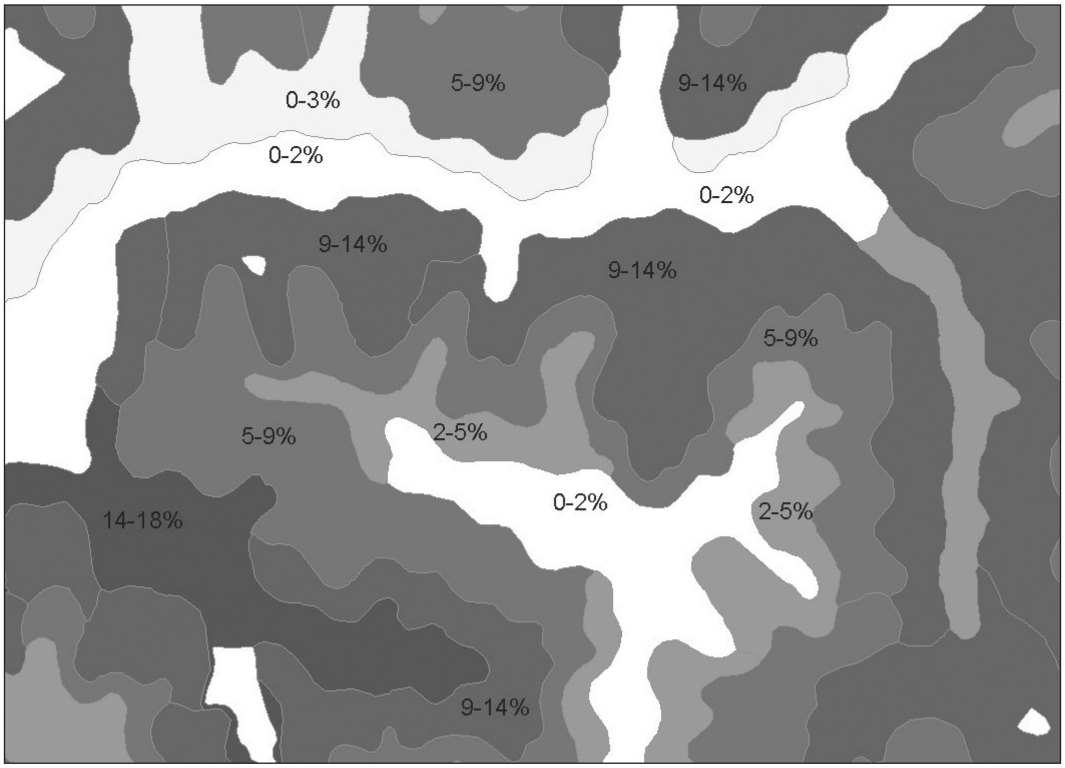


Figure 6. Soil survey soil polygons shaded by slope class range: light shades are lower slopes and darker shades indicate steeper slopes.

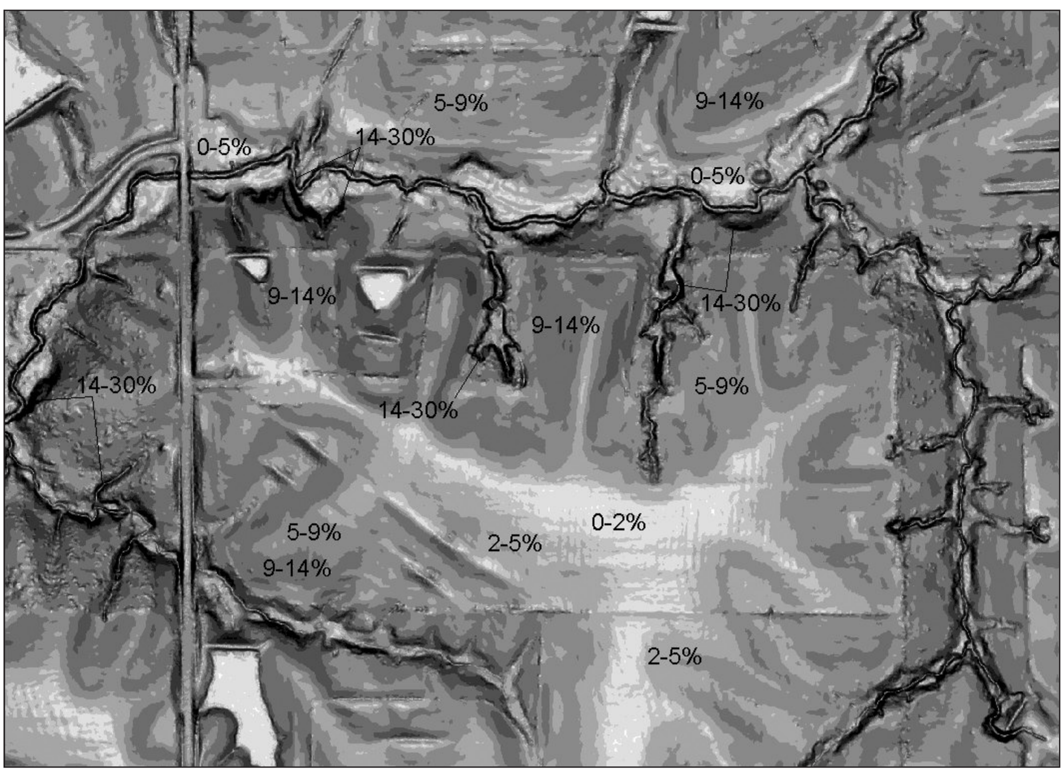


Figure 7. Slope class ranges derived from Lake Darling lidar data. While low slope areas on lidar look similar to the soil polygons, lidar shows more detail on steep slopes such as gullies and stream channels.

the near-infrared portion of the spectrum (lidar used for Lake Darling had a wavelength of 1064 nm), the intensity of lidar return is directly related to the near-infrared reflectance of the target material. An image constructed from the intensity component of the returns (Figure 8) looks very much like a black and white near-infrared aerial photograph. An intensity image has one interesting peculiarity: tree shadows point away from the flight lines, so one can see shadows pointing in opposite directions close together at the edge of two flight lines. Because intensity is recorded from each lidar return, it is possible to construct first return intensity images as well as last return intensity images, and have them look quite different. This may especially occur in forested areas where the first return might mainly represent the treetops, while the last return intensity represents many features, including the forest floor.

VERTICAL ACCURACY TEST AND INFLUENCE OF LAND COVER

Usually, one of the first questions new users of lidar have is about the vertical accuracy of the elevation data. In the Lake Darling project, the stated accuracy was 15 cm (.5') RMSE (root mean square error) in bare earth areas and 37 cm (1') in vegetated areas. Because there are no high accuracy geodetic monuments in the watershed and access to survey grade GPS equipment was unavailable, another way to test the vertical accuracy needed to be found. Fortunately, a digital terrain model and associated 2' contours produced by aerial photography and photogrammetric techniques for a road project were available from the Washington County engineer's office. This digital terrain model and contours were created by a local aerial photography firm and had a stated vertical accuracy of 6.1 cm (.2'). The area covered by the model was over 2 miles long and a quarter of a mile wide. The digital terrain model consisted of elevation points and break lines (Figure 9) in CAD format. Using the 3D_ANALYST extension in ArcGIS, the photogrammetrically derived terrain model was converted into a triangulated irregular network or TIN, and interpolated into a 1 meter elevation grid. The lidar elevation grid was subtracted from the grid made by photogrammetry to produce a simple difference grid. The overall average difference between the two grids was only 3.3 cm (.11'). To compare the two grids to their stated accuracies, the RMSE had to be calculated. First, the simple difference grid was multiplied by itself to create the squared difference grid. Using a polygon coverage of land cover from 2005, the mean squared difference was calculated for each land cover class using the zonal statistics command in ArcToolBox. By using the spatial calculator function in the SPATIAL_ANALYST extension, the square root of the values in the "mean" field

Figure 8. Portion of a lidar intensity image of the Lake Darling watershed, constructed from bare earth return intensity values.

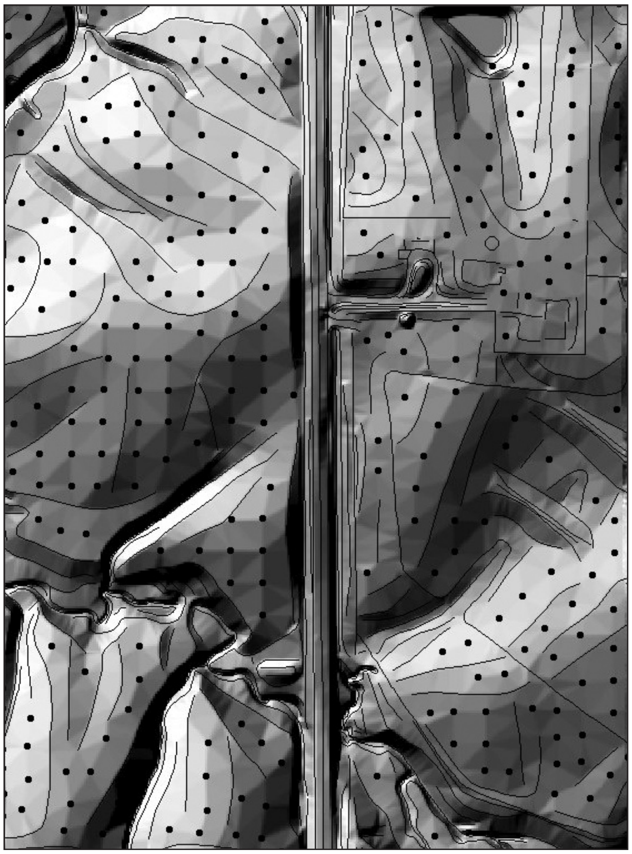


Figure 9. Portion of shaded relief digital terrain model derived from low altitude aerial photos. The black dots are elevation mass points and the black lines are break lines.

of the table was calculated, and the RMSE was found for each land cover class. The zonal statistic tool also computes a "count" of cells for each class and a "sum" of the elevations within that class. By calculating the sum of all the "count" field values and "sum" field values for all the classes, and dividing the total sum by the total count, the average squared difference for the entire dataset was found. By taking the square root of this value, the RMSE was found for the whole area. Initially, RMSE between the lidar DEM and the photogrammetry DEM was found to be .79 feet or 24.1 cm.

Upon examination of the squared difference image, it was apparent that the terrain in several areas had changed significantly between the time of the airphoto flight in 2000 and the lidar flight in 2005. These mainly included areas where the installation of sediment retention structures and dams, and road grading had occurred. When these areas were digitized and excluded from the squared difference calculation, the overall RMSE was found to be .57 feet or 17.4 cm (Figure 10). The RMSE of the row crop area was .46' (14.3 cm), grass areas .62' (18.9 cm), and forested areas .85' (25.8 cm). If the DEM derived by photogrammetric means is accepted as the higher accuracy source, then the lidar meets its stated accuracy of 15 cm in the bare ground areas, and under 37 cm in the vegetated areas. This appears to be a good test of lidar accuracy because it includes many types of land cover conditions, not just a few high accuracy locations at benchmarks on roads or nearby ditches.

COMPARING OLD AND NEW DATA

One of the first tests of any new lidar data set is to compare it with the existing DEM derived from the 10' contours from the USGS topographic quadrangle mapping projects of the latter half of the last century. Displayed at smaller scales, it is difficult to see much difference between the shaded relief images derived from the 30 meter resolution National Elevation Dataset or NED (http://ned.usgs.gov/) and lidar shaded relief. Only when the display is zoomed into larger scales is it possible to see the marked differences between the 30 meter NED (Figure 11) and lidar DEM (Figure 12). Visible on the lidar image (but not on the 30 meter NED shaded relief image), are man-made features such as roadways, ditches, fence lines, terraces and dams. Natural features such as stream channels, gullies, and floodplains are also visible.

Lidar excels at mapping topographically challenged areas: areas with little relief. Figure 13 is a shaded relief image, derived from the 30 meter NED, that shows typical glaciated terrain in north-central Iowa, east of Spirit Lake in Dickinson County. Figure 14 shows the same area using 1 meter resolution lidar, which focuses the indistinct mounds seen on the NED shaded relief into sharply defined, circular, and elongated features. These are interpreted to be remnants of ice walled lakes, which formed on the surface of the glacier. These lakes had varying amounts of sediment deposited in them, and after the ice melted, these sediments formed indistinct, low mounds (Quade et al., 2004).

Figure 15 shows the Missouri River floodplain north of Council Bluffs, Iowa, in a view, which again, uses the 30 meter resolution NED to create a shaded relief image. It reveals numerous defects in the original conversion of widely spaced contours on a very flat surface. With a 10' contour interval, there is not enough information to interpolate features on the floodplain adequately. The shaded relief image reveals cross-shaped artifacts within the DEM, which were created by the interpolation software's attempt to connect widely spaced data. Figure 16 shows the great improvement afforded by interpolating a surface from closely spaced lidar points (about 2 meter lidar postings). Missouri River meander scars, levees along drainage ditches, fence lines, interstate lanes, railroad right-of-ways, borrow pits, and sewage lagoons are all visible on the lidar shaded relief image.

When using shaded relief images for on-screen digitizing, geological mappers will need to become accustomed to recognizing and separating man-made as well

OID	LAND_USE	ZONE_C	COUNT	AREA	MIN	MAX	RANGE	MEAN	STD	SUM	RMSE_ft	RMSE_CM
0	Residential	1	36665	36665	0.00000	18.04	18.04	0.262028	0.823366	9607.27	0.511887	15.6023
1	Water	2	26069	26069	0.00000	10.053	10.053	0.234739	0.848544	6119.4	0.484499	14.7675
2	Pasture	3	152359	152359	0	41.5079	41.5079	0.383737	1.24089	58465.8	0.619465	18.8813
3	CRP	4	160257	160257	0	38.4666	38.4666	0.252299	0.855599	40432.7	0.502294	15.3099
4	Timber	5	121869	121869	0	106.069	106.069	0.71515	2.79821	87154.6	0.845665	25.7759
5	Wildlife/Wooded	6	1825	1825	0.00001	7.3607	7.36069	0.965743	1.03896	1762.48	0.982722	29.9534
6	Road	7	44609	44609	0	107.915	107.915	0.316614	0.942917	14123.9	0.562685	17.1506
7	Row Crop	8	353980	353980	0	328.666	328.666	0.214536	0.903116	75941.3	0.463180	14.1177
8	Alfalfa	9	9979	9979	0	8.07532	8.07532	0.179604	0.476496	1792.27	0.423797	12.9173
	TOTALS		907612							295399		

Mean Squ. Difference = 295399/907612 = .3255

Square root of MSD = .5705

RMSE = .57' or 17.4 cm

Figure 10. Root mean square error (RMSE) calculation of photogrammetrically derived DEM and lidar DEM, after 2000/2005 landscape-change areas removed from calculation.

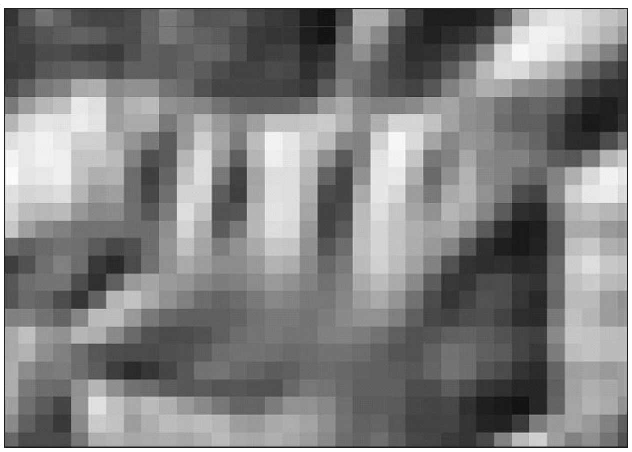


Figure 11. Portion of a shaded relief image made from a National Elevation Dataset (NED) 30 meter resolution DEM. Area is from Lake Darling watershed in Washington County, Iowa.

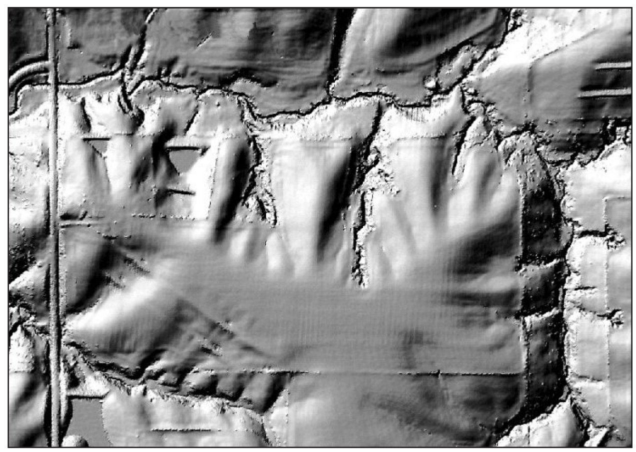


Figure 12. Portion of shaded relief image made from a 1 meter resolution lidar DEM for the same area in Washington County, Iowa.

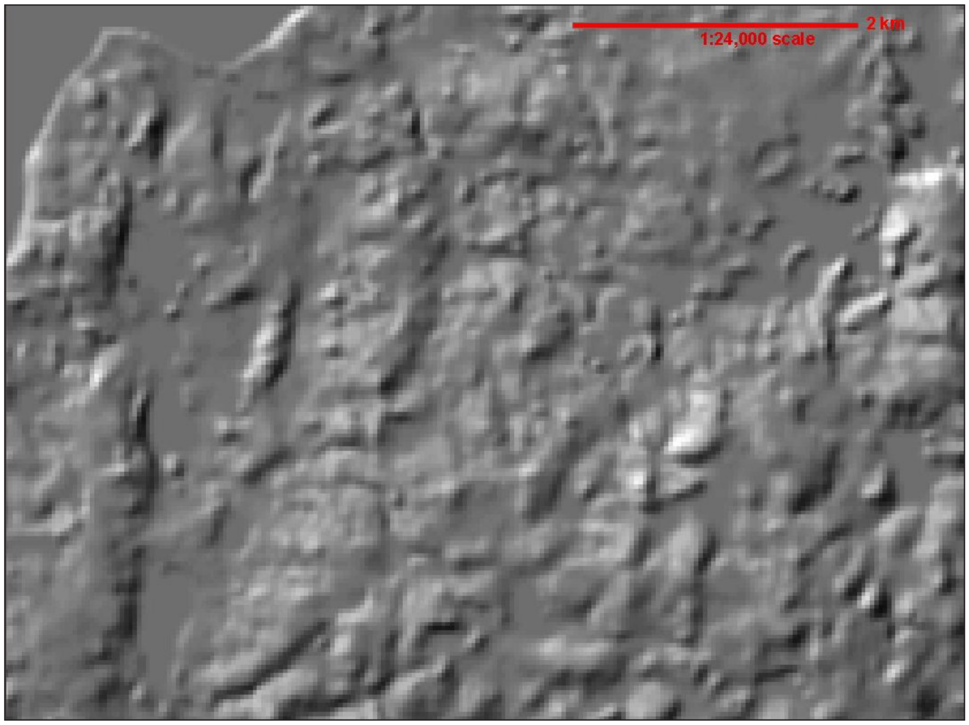


Figure 13. Portion of a shaded relief image showing recently glaciated terrain near Spirit Lake in Dickinson County, Iowa. The shaded relief was created from a 30 meter resolution DEM from the National Elevation Dataset (NED).

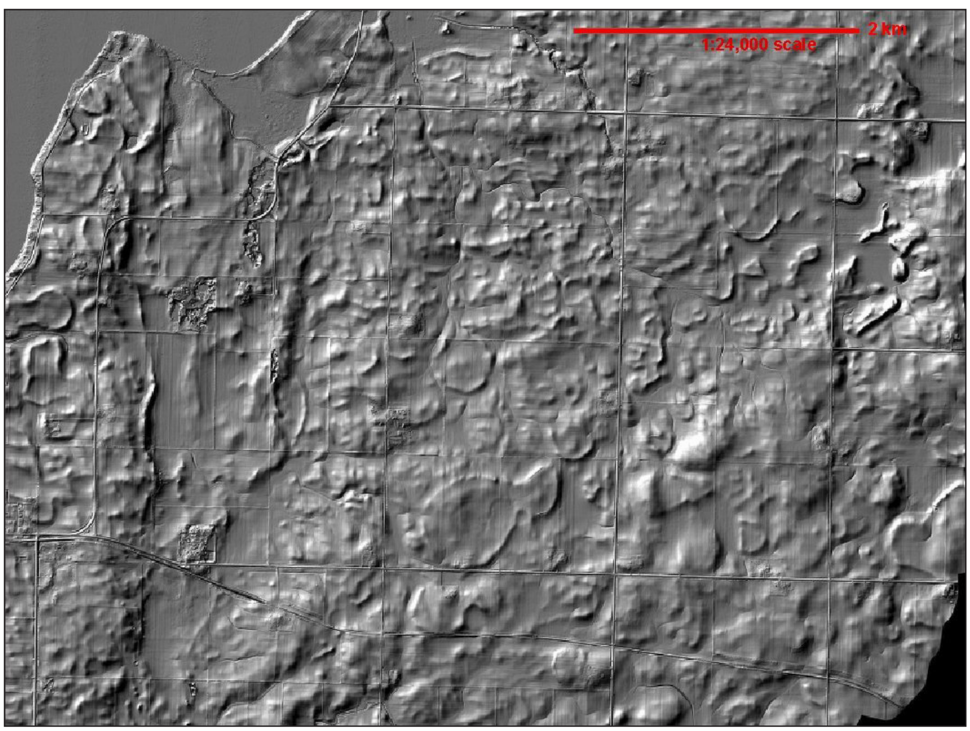


Figure 14. Portion of a lidar-derived shaded relief image of the same area of glacial terrain near Spirit Lake in Dickinson County, Iowa. Notice how the shapes of subtle, low relief glacial features are now readily apparent.

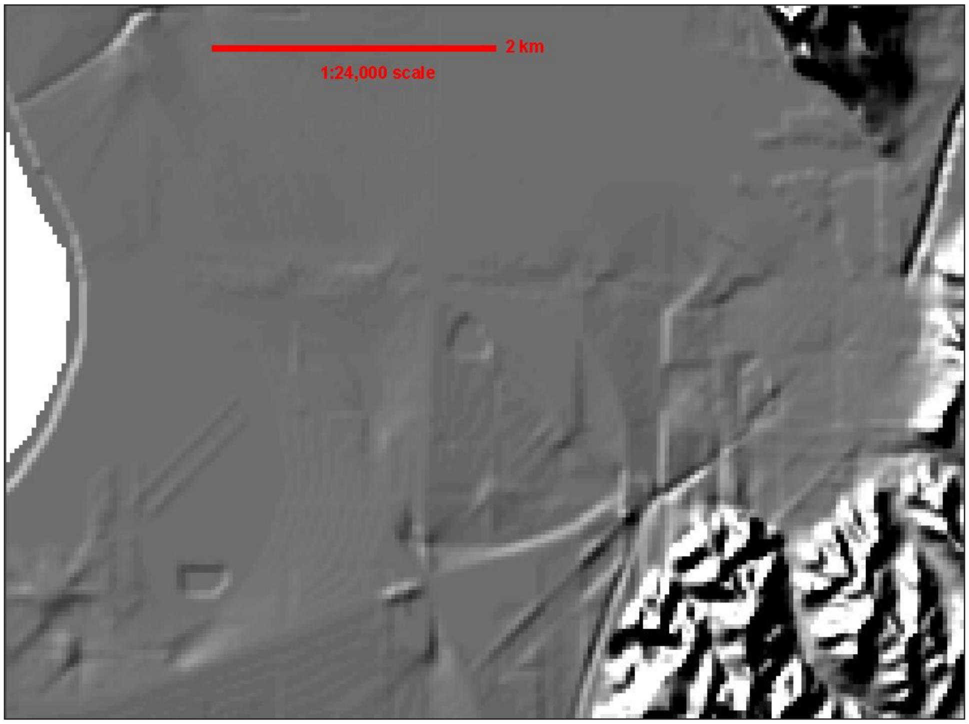


Figure 15. Portion of a shaded relief image showing the Missouri River floodplain north of Council Bluffs, Iowa. The shaded relief image was created from a 30 meter resolution DEM from the National Elevation Dataset (NED). Notice the cross-shaped features that are artifacts of the interpolation of the original 10’ contours from USGS topographic maps.

Figure 16. Portion of a lidar-derived shaded relief image of the same area on the Missouri River floodplain (see Figure 15). Notice the much finer detail showing the interstate cloverleaf, river meander scars, borrow pits, and a ditch and levee system. Lidar DEM was obtained from the Pottawattamie County GIS Department.

as geomorphic features. Because shaded relief images can represent the encoding of relatively small changes in slopes, mappers will need to build up criteria for recognition of everyday features using the clues in contrast, shading, shape, texture, pattern, and context contained in these images. In the past, geological mappers learned how to interpret aerial photos by poring over example after example of natural and man-made features. They also learned how to interpret geological features by looking at geomorphic signatures on topographic maps. Lidar will require relearning and reinventing both techniques by moving the geomorphic scale down to the realm of the airphoto, roughly at resolutions from 1 to 5 meters. While qualitative information on slopes was available by way of stereo viewers and aerial photos, there has never been as much quantitative slope information available until now with the advent of lidar data. With digital elevation data derived from lidar, new computer assisted classification strategies for geomorphic feature interpretation can be developed, as can new types of imagery to support manual interpretations.

SUMMARY

Large-scale lidar acquisitions will provide mapping professionals with an increase of new, high quality elevation data to use as base maps for their projects. To take full advantage of this new data source, those who are mapping need to be aware of how lidar data are collected and what data reduction processes commercial vendors use to make deliverable products for their clients. In many cases, mappers will want to manipulate the raw lidar returns into their own TINs, DEMs, and derived products, but sometimes they will only have access to vendor-supplied, finished products that have undergone unknown procedures to make the visual appearance more appealing. Mappers can use shaded relief images derived from lidar DEMs or TINs for on-screen digitizing, as well as new derivative products such as terrain slope and lidar intensity to identify geologic features and other features. Anyone using lidar data will be interested in the absolute vertical accuracy of elevations and will need to know how land cover type affects that accuracy.

REFERENCES

Baltsavias, E.P., 1999, Airborne laser scanning: basic relations and formulas: ISPRS Journal of Photogrammetry and Remote Sensing, no. 54, p. 199-214.

Quade, D.J., Giglierano, J.D. and Bettis, III, E.A., 2004, Surficial geologic materials of the Des Moines Lobe of Iowa, Phase 6: Dickinson and Emmett Counties: IDNR/IGS, OFM-0402, 1:100,000 scale, accessed at http://www.igsb.uiowa.edu/gsbpubs/pdf/ofm-2004-2.pdf.

Wehr, A. and Lohr, U., 1999, Airborne laser scanning – an introduction and overview: ISPRS Journal of Photogrammetry and Remote Sensing, no. 54, p. 68-82.

Geological Map Database – A Practitioner's Guide to Delivering the Information

By Jeremy R A Giles

British Geological Survey
Kingsley Dunham Centre
Keyworth
Nottingham, United Kingdom
NG12 5GG
Telephone: +44(0)115 936 3100
Fax: +44(0)115 936 3200
e-mail: jrag@bgs.ac.uk

INTRODUCTION

The British Geological Survey (BGS) has been practicing geological mapping since 1835. One would think that we should be getting quite good at it by now. The simplistic view is that the geology doesn't change, or at least not very quickly, so why is the job not done? There are two related answers to that question. The first is that our understanding is continually improving, so that we can know more about any given area. The second is that the demands upon the outputs of geological surveying are ever increasing. In his Presidential Address to the Geological Society of London in 1836, Sir Charles Lyell explained the process of setting up the world's first national geological survey "to cover the cost of geologically coloring the topographical maps of the trigonometrical Survey." He said: "…we drew up a joint report in which we endeavoured to state fully our opinion as to the great advantages which must accrue from such an undertaking not only as calculated to promote geological science, which would alone be a sufficient objective, but also as a work of great practical utility bearing on agriculture, mining, road-making, the formation of canals and railroads and other branches of national industry".

Those demands have now grown considerably as the number and variety of the branches of national industry has grown and developed. Roger Tym & Partners estimated in November 2003 that: "the total value added of national output to which BGS contributes for 2001 lies in the range of $64 billion – $116 billion, representing around 5%–8% of total UK output (GVA). This is of course orders of magnitude greater than BGS's annual turnover of approximately $75 million."

The aims of BGS geological mapping are stated in Walton and Lee (2001). They said: "the key objectives of the programme are to (i) deliver high quality detailed information on bedrock and superficial geology of the UK landmass as digital, map-based and text data, (ii) provide increased information on Quaternary and other superficial deposits, (iii) provide increased understanding of three-dimensional structure and process, (iv) deliver the near-surface component of the Digital Geoscience Spatial Model (DGSM) and (v) deliver the remaining 'sheets' in the current programme for incorporation into the Digital Geological Map of Great Britain (DiGMapGB)."

To achieve these stated aims, BGS has been progressively developing the key components of a geological mapping system (Figure 1). This multi-component system has been developed by a large team of scientists and developers (see Acknowledgements). The key components of the system, which are described below, are as follows:

- BGS-geoIDS and the associated databases and data stores in which BGS information is managed
- NGDC
- SIGMA
- LithoFrame
- Internal Discovery
- Internet
- DiGMapGB
- GeoSure

BGS GEOSCIENCE INTEGRATED DATABASE SYSTEM (BGS-GEOIDS)

BGS is the custodian of a wealth of geoscience information that has been collected by its own scientists or deposited by industry under various government statutes and voluntary agreement. Acquisition of this information has been continuous since the Survey's formation in 1835, and material created prior to this date is also stored in its archives. The range of information types includes materials (such as rocks, fossils, minerals, and borehole core), paper records, microfiche, reports, digital databases, digi-

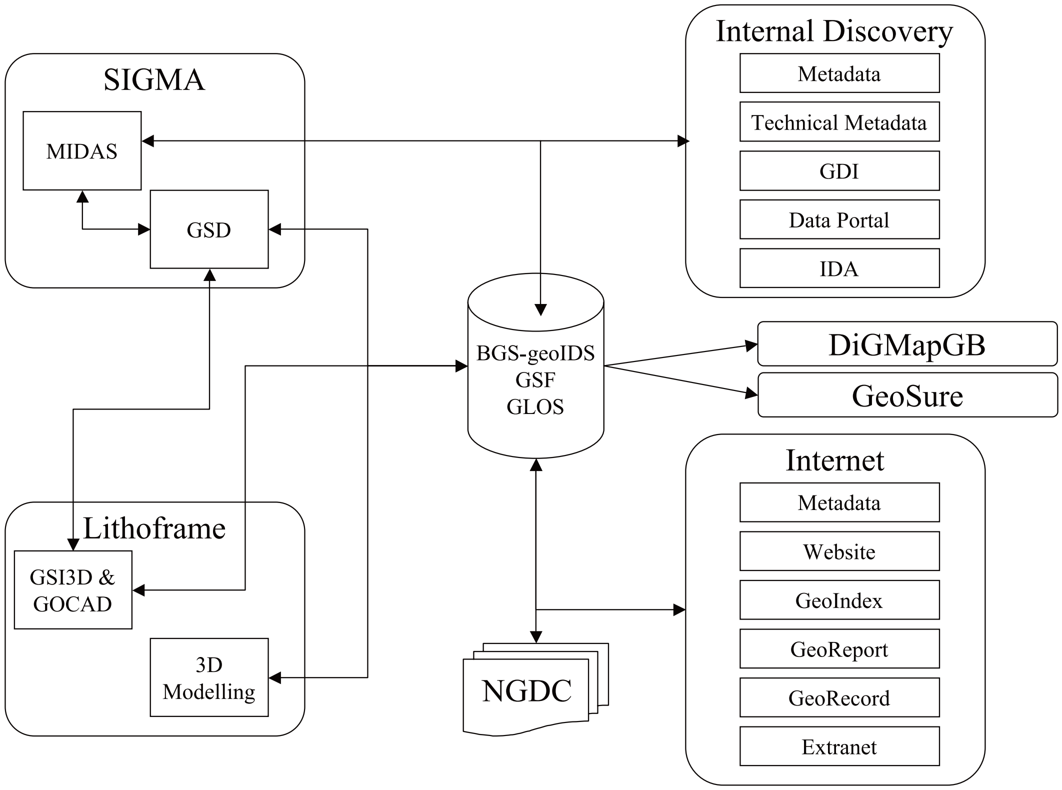


Figure 1. Geological Mapping Implementation (BGS-geoIDS – BGS Geoscience Integrated Database System; GDI – Geoscience Data Index; GLOS – Geoscience Large Object Store; GSD – Geoscience Spatial Database; GSF – Geoscience Spatial Framework; GSI3D – Geological Surveying and Investigation in 3D; IDA – Intranet Data Access; MIDAS – Mobile Integrated Data Acquisition System; NGDC – National Geoscience Data Centre; SIGMA – System for Integrated Geospatial Mapping).

tal files, digital models, etc. A range of systems and tools has been developed to manage these information assets in a holistic manner.

Peebler (1996) made the following observation: "Lack of basic data integration costs the average E&P professional a considerable amount of time. According to various estimates geoscientists and engineers spend from 20% to 30% of their total project time searching for, loading and formatting data."

Similarly, Adam Dobson (*Pers Comm*: 2002), representing Shell, a major international oil company, said that an internal audit undertaken in 2002 showed that, for new frontiers areas, staff spent their time as follows:

- Finding data – 53%
- Archiving data – 9%
- Documenting the data – 15%
- Interpreting (adding value) – 23%

On the basis of this audit Shell set a target of reducing the time spent finding data to 30% and increasing the adding-value time to 46%.

Several years earlier the BGS-geoIDS Project had been established to resolve a number of similar problems. BGS recognized that it held a wealth of valuable and important data, but that this resource was largely inaccessible to staff. Some of the data might be held and managed in well designed databases, but these were isolated "islands of excellence." There was little interoperability between these islands, so that routine integration and onward use or enhancement of the data were rarely straightforward. There were few corporate standards, and the local standards that did exist were not shared between databanks. BGS had no maintained metadata, so most BGS staff had no idea what data were held corporately, how these data might be used, or what their quality was. Finally, there were no corporate application standards, so data were accessed through the use of a multitude of different tools that had been built with no consistent design standards and no thought for future interoperability.

Key drivers recognized in planning and undertaking the BGS-geoIDS work were the opportunities to

- reduce staff effort in finding data,

- make quality-assured data available to staff and customers,
- encourage and facilitate collaboration across BGS,
- improve access to the unique BGS information base,
- keep BGS at the forefront of the development of digital geoscience systems,
- inform and support management decisions,
- create and implement corporate standards and establish best practice

The BGS-geoIDS Project produced a range of deliverables, including a Corporate Data Policy, a system for data management planning, metadata at various levels, a documented corporate data model, an application standard implemented through an Intranet data access tool, and the adoption of BGS-wide best practice. Above all, the Project imposed a significant culture change within BGS, transforming data from being personal property to being corporate property.

Subsequently, the DGSM (Digital Geoscience Spatial Model) Project extended the data management system to deal with digital 3D models and introduced the Geoscience Large Object Store (GLOS) to hold the full models in their proprietary software format with associated metadata. However, it was recognized that the various proprietary software formats were unlikely to remain unchanged and could not safely be used for archiving the models. Such models would probably have a life expectancy of less than 10 years. Thus, a second component was introduced, which "sampled" the model and produced a series of X, Y, Z coordinates for each stratigraphical horizon represented in the model. As this information is stored as a simple digital file, it is more suitable for long-term preservation.

NATIONAL GEOSCIENCE DATA CENTRE (NGDC)

www.bgs.ac.uk/ngdc

All BGS corporate data are managed through the National Geoscience Data Centre, which is the Natural Environment Research Council (NERC) designated center for geoscience data and information management, and which has five main elements:

- National Geoscience Records Centre
- National Geoscience Materials Collection
- National Hydrocarbons Data Archive
- NGDC Earth Science Academic Archive
- NGDC Digital Data Management

The top level aim of the NGDC is to manage all BGS data and information in accordance with the NERC and BGS Data Policies. NGDC staff members manage a wide range of information types, aiming to preserve them for use by future generations. Thus, the environments in which the collections are held are monitored carefully, and action is taken to manage environments where conditions fall outside accepted norms. For example, localised high humidity in a room holding a palaeontological collection triggered an investigation that discovered a fractured rain waste pipe on the exterior of the building. A robust metadata system is recognized as being indispensable, and an active program of metadata management is operated within the NGDC. Where appropriate, digital indexes are created and maintained as aids for finding individual records or specimens.

Overall, the NGDC activities attempt to strengthen users' confidence by creating and maintaining validated and verified datasets to agreed standards, and providing tools that enable geoscientists and others who need geoscientific insight, both inside and outside BGS, to use BGS information with confidence.

SYSTEM FOR INTEGRATED GEOSPATIAL MAPPING (SIGMA)

http://www.bgs.ac.uk/scripts/downloads/start.cfm?id=381

Once the key elements of information that underpin geological mapping are in place and tools have been provided to facilitate access, it becomes possible to build a digital geological workflow that starts with digital field capture, progresses through digital map compilation, and passes into a digital map production and management system. The project that specified and developed this process is called SIGMA. The project has two key elements, the first being the MIDAS system (Mobile Integrated Data Acquisition System), the second the Geoscience Spatial Database (GSD).

Mobile Integrated Data Acquisition System (MIDAS)

Designed to allow use in the field, each MIDAS setup is mounted on a weather proof, robust, impact-resistant computer. It is based on standard ESRI software that has been customized to meet the project's specific requirements. The field geologist's base map and an analysis of existing information are loaded onto the computer, and the GIS provides digital field slip functionality. A global positioning system is used to locate the sites of observations, and forms are called up to support the population of a Microsoft Access Database with a range of information.

Geoscience Spatial Database (GSD)

Once the geologist returns from the field, the GSD is used to compile the geological map from existing information and from the data captured by the MIDAS system. The product is a traditional geological standard map that is created and managed in the digital environment along with its accompanying digital databases. This suite of information is then passed into the cartographic map production system for final delivery in the form of the DiGMapGB product or as printed maps.

LITHOFRAME

http://www.bgs.ac.uk/scripts/downloads/start.cfm?id=535

Merger of aspects of the data captured by the digital geological mapping workflow with appropriate digital information managed by the NGDC permits the creation of three-dimensional models of the geology of the whole or part of Great Britain. For example, the BGS receives approximately 50,000 borehole logs a year from industry. These range from shallow construction industry boreholes to deep energy exploration boreholes. The borehole logs are scanned and the metadata entered into the appropriate database by the NGDC registration team. The borehole logs are then available for use, on every desktop in the BGS, and can be accessed through a range of application and GIS tools. Two principal tools are used to undertake the modeling. The first is GSI3D, which is used for modeling superficial deposits, and the second is GoCAD, which is used for bedrock modeling. LithoFrame models are prepared at various resolutions:

- LithoFrame – shows the most significant stratigraphical divisions and major faults
- LithoFrame250 – prepared for stratigraphical groups
- LithoFrame50 – modelled at the formation level
- LithoFrame10 – focuses on well-characterized and relatively shallow superficial deposits

Geological Surveying and Investigation in 3D (GSI-3D)

http://www.bgs.ac.uk/science/3Dmodelling/gsi3d.html

The GSI-3D software tool and methodology has been developed over the last decade by Dr Hans-Georg Sobisch of INSIGHT Geological Software Systems GmbH, based in Cologne. During the past 3 years, BGS has acted as a test bed for the accelerated development of the tool and methodology. GSI-3D utilizes a Digital Terrain Model, surface geological linework, and downhole borehole data to enable the geologist to construct cross sections by correlating boreholes and the outcrops to produce a geological fence diagram. Mathematical interpolation between the nodes along the drawn sections and the limits of the units produces a solid model comprising a stack of triangulated objects, each of which correspond to one of the geological units present. Geologists draw their sections based on facts such as borehole logs correlated by intuition – the shape "looks right" to a geologist. This "looks right" element draws on each geologist's wealth of understanding of earth processes, examination of exposures, and theoretical knowledge gathered during a career in geology.

GoCAD

http://www.gocad.org/www/

GoCAD is the tool used in BGS for modeling bedrock geology, as it has additional features, such as fault and fold handling capabilities, which GSI-3D lacks. The GoCAD Research Program is run by the Computer Science Group of the National School of Geology in Nancy, France. This project is currently undertaken in collaboration with the Institut National Polytechnique de Lorraine and the Centre de Recherches Petrographiques et Geochimiques, France. The aim of the research program is to develop a new computer-aided approach for the modeling of geological objects. This approach is specifically adapted to geophysical, geological, and reservoir engineering applications.

KNOWLEDGE DELIVERY

The knowledge created during the geological mapping process is delivered through a range of products and services that are available from the BGS (see "Internal Discovery" and "Internet", in Figure 1). These include not only the geological map itself but also the elements of data that were used to develop the map. These products include:

- The BGS Website
- The BGS Intranet
- Discovery Metadata
- GeoIndex
- GeoReport
- GeoRecord
- DiGMapGB
- GeoSure

Website

http://www.bgs.ac.uk/

The award winning BGS website is accessed extensively by a wide range of users. It caters specifically to the needs of various groups, ranging from children to professional geoscientists. Its aim is to inform users about BGS activities and provide them with access to information and data. The site offers well over 900 downloads, a number set to rise to well over 10,000 in the near future. There is access to a range of definitive data sources. For example the four BGS Rock Classification Scheme (RCS) reports are available from http://www.bgs.ac.uk/bgsrcs/home.html. The BGS Rock Classification Database can be accessed and searched using a web-based form at http://www.bgs.ac.uk/bgsrcs/searchrcs.html, and the data that it holds can be downloaded in spreadsheet format at http://www.bgs.ac.uk/data/dictionaries.html.

Intranet

In parallel to the BGS Website, there is a comprehensive Intranet that provides information and data to BGS staff. Intranet information and data suites are normally more comprehensive than their Internet equivalents.

Discovery Metadata

www.bgs.ac.uk/metadata

Published BGS discovery metadata can be accessed as part of the main BGS Website. A profile of ISO19115 is used to describe each dataset at a level that is appropriate for a user to assess whether the data contained in the dataset are appropriate for their needs and to allow recognition of data limitations. The full record is posted on the Internet, and the tools to manage the underpinning Oracle database are on the Intranet.

Technical Metadata

The BGS Technical Metadata, which describe the numerous components to the BGS Oracle database system, are only available on the Intranet. The information provided includes details of databanks, tables, views, indexes, synonyms, etc., and the system as a whole contains some of BGS's most critical digital data.

As the Technical Metadata system is complex, it is actively maintained to help users. The system extends Oracle's own data dictionary and is designed to help those with a basic understanding of Oracle to navigate the objects that make up the BGS Data Architecture. Its application front-end also provides "Best Practice Guidelines for Oracle Development," procedures for changing the structure of database objects and some documentation on data models.

GeoIndex

http://www.bgs.ac.uk/geoindex/home.html

The GeoIndex provides detailed metadata about selected datasets. Using a web-based GIS, it shows the locations of data points within each featured dataset and provides basic information about each data point. For example, the sites of geochemical stream sediment samples are shown in the GIS along with a list of elements analyzed in each sample.

Geoscience Data Index (GDI)

The intranet version of the GeoIndex is the Geoscience Data Index (GDI). It is built in ESRI's ArcGIS and allows BGS staff to discover the availability of spatially-referenced information, drill down to it, and gain access to it. It allows a rapid assessment to be made of information and data that are available for any given location within Great Britain.

Data Portal

The BGS Data Portal is an Intranet tool that allows access to the datasets and information to support 3D modeling. Its functionality overlaps partially with that of the GDI and the two applications will eventually be merged.

Intranet Data Access

Intranet Data Access (IDA) is an Intranet tool available to BGS staff to facilitate access to BGS Oracle databases through a user-friendly forms interface. The interface has been developed using Adobe's ColdFusion, and its components have been developed to a standard design template so that users are presented with a consistent look and feel across the application. Examples of the databases that can be accessed include:

- Borehole locations
- Borehole lithology
- Borehole materials (samples collected from boreholes)
- Geophysical log index
- Geological maps and field slips
- Palaeosaurus (paleontology specimen data)
- Britrocks (rock specimens)

GeoReport

http://shop.bgs.ac.uk/georeports/

If you are investigating land or property, GeoReports could save you time and money. GeoReports will:

- tell you about the condition of the ground—its geology, hydrogeology and any related hazards (such as subsidence or radon potential),
- let you know what information about your site might already be held in the national geological archive,
- provide cost-effective access to expert advice from BGS scientists who know about your local area.

There are a wide range of GeoReports that are available from the BGS Internet site, including:

- Building stone assessments
- Data listings
- Geological assessments
- Geological map extracts
- Ground source heat pump
- Ground stability
- Radon protection
- Water borehole prognosis

GeoRecord

http://www.bgs.ac.uk/boreholes/home.html

The BGS provides a comprehensive range of scanned documents, including:

- Borehole records
- Site investigation reports
- Technical reports
- Mine records

DIGMAPGB

http://www.bgs.ac.uk/products/digitalmaps/home.html

BGS maps are increasingly being offered digitally, either as raster images or as vector data, in a variety of formats and structured into themes. This allows them to be used in Geographical Information Systems (GIS) where they can be integrated with other types of spatial data to provide powerful aids to problem solving in many earth science situations. Map data are available at a range of resolutions, from small to large scale, and they cover many aspects of the geological and related sciences.

Layers include:

- Superficial deposits
- Artificial ground
- Mass movement
- Thickness of superficial deposits
- Bedrock geology
- Elevation of bedrock

GEOSURE

http://www.bgs.ac.uk/products/geosure/home.html

Hazards that go unrecognized by developers, householders, or local government may lead to financial loss, which can be avoided. The cost of arresting or repairing a ground failure is far greater than the cost of prevention.

Understanding geology is vital when determining the stability (and thus the value) of land and property, and ensuring the safety of its occupiers. The GeoSure datasets from the British Geological Survey provide information about potential ground movement or subsidence in a helpful and user-friendly format. The datasets can help inform planning decisions and indicate potential causes of subsidence:

- Soluble rocks (dissolution)
- Shrink-swell clays
- Landslides (slope instability)
- Compressible ground
- Running sands
- Collapsible deposits

As well as being able to license any of these datasets in digital form, the BGS provides a report generation service GeoReport, whereby reports can be produced giving details of six ground stability issues for specified areas or properties.

CONCLUSION

The BGS has aspired to develop a digital workflow for its data and information acquisition, management, manipulation, and delivery. This aspiration is now approaching fulfillment. Tools that have been developed will allow digital capture of field mapping data and support its enhancement right through to its digital delivery.

Digital information created by the process is being managed systematically across the entire organization in a manner that allows its rapid discovery, retrieval, and exploitation.

ACKNOWLEDGEMENTS

The BGS multi-component geological mapping system has been built by the efforts of a large team of people including:

- Keith Adlam, GIS Expert
- Garry Baker, Deputy Programme Manager for Information Management
- Brian Cannel, Webmaster
- Bill Hatton, Programme Manager for Information Systems Development
- Andrew Howard, Programme Manager for National Geoscience Framework

- Richard Hughes, Programme Manager for Information Delivery
- Ian Jackson, Programme Director for Information
- David Lowe, Geoscience Data Manager
- Hans-Georg Sobisch of INSIGHT Geological Software Systems
- Martin Nayembil, Data Architect
- Rob Pedley, Application Architect
- Jenny Walsby, Programme Manager for Information Products

REFERENCES

Peebler, R., 1996, Extended Integration: The Key To Future Productivity Leap, *Oil and Gas Journal* May 20, 1996, Vol.94. No.21.

Tym, R., 2003, The Economic Benefit of the BGS: British Geological Survey, 8pp, accessed at http://www.bgs.ac.uk/about/economicbenefits.html.

Walton, G and Lee, M. K., 2001, Geology for our diverse economy: report of the Programme Development Group for Onshore Geological Surveys: Keyworth, Nottingham, British Geological Survey, 99p.

Preservation of North Carolina Legacy Geologic and Topographic Maps: A Cooperative Effort with the North Carolina Geological Survey, North Carolina State University and the Library of Congress

By Jeffrey C. Reid[1], Jefferson F. Essic[2], Steven P. Morris[2], Smitha Ramakrishnan[2], and Julia L. Harrell[3]

[1]North Carolina Geological Survey
1612 Mail Service Center
Raleigh, NC 27699-1612
Telephone: (919) 733-2423
Fax: (919) 733-0900
e-mail: jeff.reid@ncmail.net

[2]North Carolina State University, Raleigh, D.H. Hill Library
e-mail: {jeff_essic, steven_morris, snramakr}@ncsu.edu

[3]North Carolina Department of Environment and Natural Resources
e-mail: Julia.Harrell@ncmail.net

Geologic and historic maps are in high demand, and are critical for earth science instruction and research. Digital georeferencing of scanned maps provides new capabilities not possible with paper maps by allowing other data to be overlain by, and analyzed with, these map images using GIS software. Therefore, geologic and topographic maps for North Carolina are being collected, scanned, georeferenced, and preserved in a collaboration between the North Carolina Geological Survey (NCGS) and North Carolina State University Libraries (NCSU Libraries). The legacy geologic and topographic maps have no digital counterparts, and paper copies are scarcely accessible.

Geologic maps, including collars, are scanned with a large format (42-inch) HP Designjet 815 mfp scanner-plotter device to create a 300 dpi TIF file. Using ArcMap 9.1's georeferencing extension, at least four geographic locations were interactively selected from the TIF, based on coordinates and grids printed on the maps. ArcMap creates a table of these selected coordinate values, and with U.S. Army Corps of Engineers Corpscon software, each coordinate pair is converted to NC Stateplane NAD83 meters and then appended to the table. ArcMap then creates a TFW world file and transforms the TIF image so that it is represented in the data view in Stateplane Coordinates (NAD83). Each image file is rectified with ArcInfo workstation and compressed with MrSID. The workflow is shown in Reid et al. (2006a), Reid et al. (2006c), and Essic et al. (2006); Ramakrishnan (2006) also provided details. Previously, Cahill et al. (2002) reported on the scanning and delivery of historic maps over the Internet as done by the Library of Congress.

The TIF images and world files will become part of the North Carolina Geospatial Data Archiving Project, a partnership between NCSU Libraries and the Library of Congress (North Carolina Geospatial Data Archiving Project, http://www.lib.ncsu.edu/ncgdap). As of June, 2006, the inventory consists of 101 U.S. Geological Survey geologic maps, 130 North Carolina Geological Survey geologic maps, 47 maps from theses and dissertations, 8 N.C. Department of Transportation maps, and 165 legacy 15-minute topographic maps, all 451 of which are backed up on multiple secure servers (http://wfs.enr.state.nc.us/NCGeologicMaps/; Reid et al. (2006d)).

Data are planned for dissemination through the North Carolina Geological Survey's Geologic Map Catalog, (http://wfs.enr.state.nc.us/NCGeologicMaps/), the NCSU Library's campus-wide server (http://www.lib.ncsu.edu/gis/geolmaps.html) pointing at the NCGS' Geologic map catalog URL, and by contribution to NCOneMap (http://www.nconemap.com) and the National Geologic Map Database's Map Catalog (http://ngmdb.usgs.gov/).

The compressed MrSID geologic and topographic maps accompanied by their world files and supplemental data (MS Excel and ESRI shape files) are in Reid et al. (2006d). The legacy topographic maps are now online at: http://www.lib.ncsu.edu/gis/historictopos.html in JPEG 2000 format. Other geologic maps to fill in geographic gaps are actively sought. An annual map service and data listing update is planned.

REFERENCES

Cahill, C. R., Roberts, G. E., and Shug-O’Neill, D., 2002, Scanning and Delivery of Historic Maps Over the Web: The Library of Congress Experience, *in* Soller, David R., ed., Digital Mapping Techniques ’02—Workshop Proceedings: U.S. Geological Survey Open-File Report 02-370, pp. 99-104, available at http://pubs.usgs.gov/of/2002/of02-370/cahill.html.

Essic, J., Reid, J. C., Morris, S., and Ramakrishnan, S., 2006, Preserving North Carolina Legacy Geologic and Topographic Maps, abstract, Geological Society of America, Abstracts with Programs, Vol. 38, p. 30.

NCOneMap, accessed 16 June 2006 at http://www.nconemap.com.

North Carolina Geological Survey’s Geologic Map Catalog, accessed 16 June 2006 at http://149.168.30.180/NCGeologicMaps.

North Carolina Geospatial Data Archiving Project, accessed 16 June 2006 at http://www.lib.ncsu.edu/ncgdap.

Ramakrishnan, S, 2006, Georegistration for digital archiving and distribution of scanned geologic maps of North Carolina: A cooperative effort with the Library of Congress and NCSU Libraries, unpublished master’s degree thesis, The University of North Carolina, Greensboro, North Carolina, 35p.

Reid, J. C., Essic, J. F., Morris, S. P., Ramakrishnan, S., and Harrell, J. L., 2006a, Preserving North Carolina Legacy Geologic and Topographic Maps, MS PowerPoint, accessed at http://ngmdb.usgs.gov/Info/dmt/DMT06presentations.html.

Reid, J. C., Essic, J. F., Morris, S. P., Ramakrishnan, S., and Harrell, J. L., 2006b, Preserving North Carolina Legacy Geologic and Topographic Maps, poster part A, accessed 29 June 2006 at http://ngmdb.usgs.gov/Info/dmt/DMT06presentations.html.

Reid, J. C., Essic, J. F., Morris, S. P., Ramakrishnan, S., and Harrell, J. L., 2006c, Preserving North Carolina Legacy Geologic and Topographic Maps, poster part B, accessed 29 June 2006 at http://ngmdb.usgs.gov/Info/dmt/DMT06presentations.html.

Reid, J. C., Essic, J. F., Morris, S. P., Ramakrishnan, S., and Harrell, J. L., 2006d, Preserving North Carolina Legacy Geologic and Topographic Maps: North Carolina Geological Survey, Open-file report 2006-01, DVD (also contains Reid et al 2006a—c).

Saturation And Value Modulation (SVM): A New Method For Integrating Color And Grayscale Imagery

By David W. Viljoen[1] and Jeff R. Harris[2]

Geological Survey of Canada
615 Booth St.
Ottawa, ON, K1A 0E9
Telephone: (613) 995-1207
Fax: (613) 995-9273
e-mail: viljoen@nrcan.gc.ca; harris@nrcan.gc.ca

ABSTRACT

Algorithms for integrating color imagery with grayscale imagery have long been an important feature of many remote sensing (RS) image analysis and geographic information systems (GIS). Traditional methods for data integration include Red-Green-Blue (RGB) /Hue-Saturation-Value (HSV) transformation and RGB modulation. However, these techniques are either inflexible or present a compromise between the quality of the color and the contribution of the shading. Furthermore, these techniques can also result in serious color distortions. Layer transparency is another popular technique for integrating data that is available in most RS and GIS software packages. However, optimal integration of color and grayscale imagery is difficult to achieve using this method.

We briefly review the shortcomings of these traditional image integration methods and introduce a new method (Saturation-Value-Modulation [SVM]) for raster image integration developed by David Viljoen at the Geological Survey of Canada. SVM is flexible and does not compromise the color or grayscale components of the resulting integrated image. The general concepts behind this algorithm as well as the five parameters used to control the integration process are discussed. Various examples of how SVM can be used to integrate various geoscience data are also presented. Finally, we provide a brief overview of the ESRI ArcGIS implementation of SVM, though we do not include a detailed presentation of the actual Visual Basic code or the algorithm.

The ArcGIS map document (MXD) that contains the VBA (Visual Basic for Applications) code is available for download for those who wish to use SVM.

INTRODUCTION

There are two primary reasons for integrating a color image with a grayscale image. The first is to provide visual enhancement of a single dataset by combining different characteristics. For example, a color image of a digital elevation model (DEM) can be integrated with a grayscale image of the shaded relief DEM (Figure 1). The second is to visualize the relationship between two very different types of data. For example, gamma ray spectrometer data can be combined with Landsat Thematic Mapper band 7 (Figure 2). Many methods have been developed to integrate imagery in remote sensing image analysis and geographic information systems. It is instructive to review a few of these methods to appreciate some of the advantages of the SVM method.

Modern remote sensing software and GIS often have a layer transparency feature that facilitates the integration of data and allows the user to increase or decrease the transparency of one layer to reveal the layer that would otherwise be hidden. The advantage of this method is that it is instantaneous, as it does not involve pixel-by-pixel computations and color transformations. This method can be used with two or more color images or a color image and a grayscale image. The resulting integrated image is a weighted interpolation of the colors of the contributing images. The disadvantage of this method in integrating color and grayscale imagery is that the resulting integrated image compromises either the color or the shading (Figure 3).

Remote sensing and GIS software often have tools for performing transformations between Red-Green-Blue (RGB) and Hue-Saturation-Value (HSV) color models. Figure 4 graphically shows the components of the HSV model where hue is the dominant wavelength of the color, saturation is the presence or absence of color, and value is the brightness and darkness. Color transformations involve pixel-by-pixel conversion of RGB color components into equivalent HSV components. Integration of color and grayscale imagery is achieved by replacing the value component (V) with the values from the grayscale image (Figure 5). One of the problems with this technique is that the value component is often important in defining colors in the color image (Harris et al. 1990, 1994). That

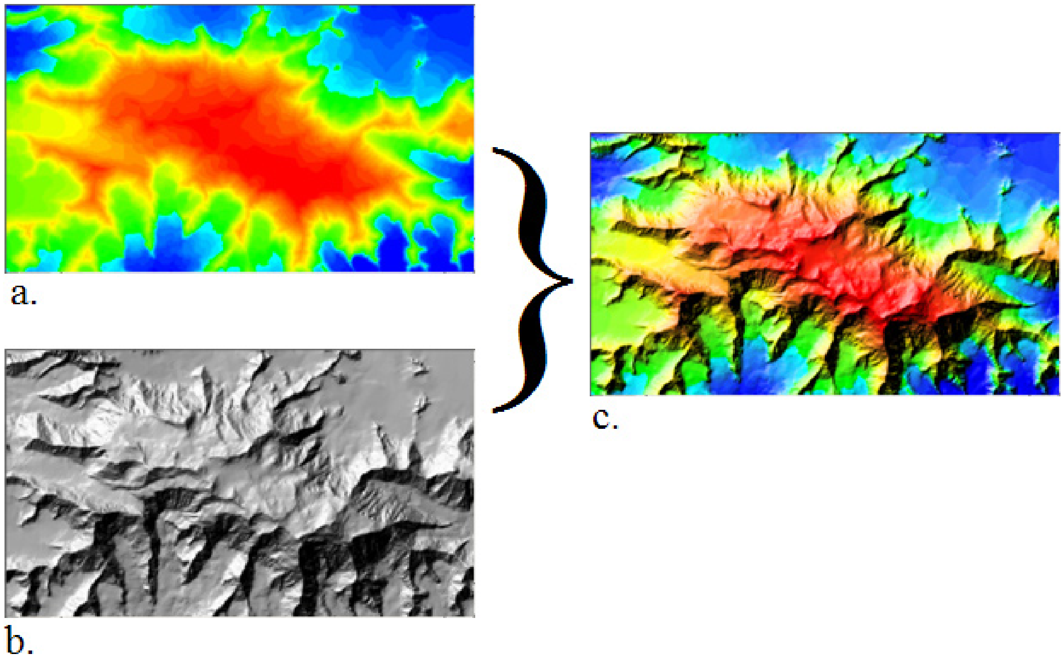


Figure 1. a. Color image of a digital elevation model (DEM) of Mt. Logan. b. Shaded relief of Mt. Logan DEM. c. Integrated image using Saturation-Value-Modulation (SVM) method.

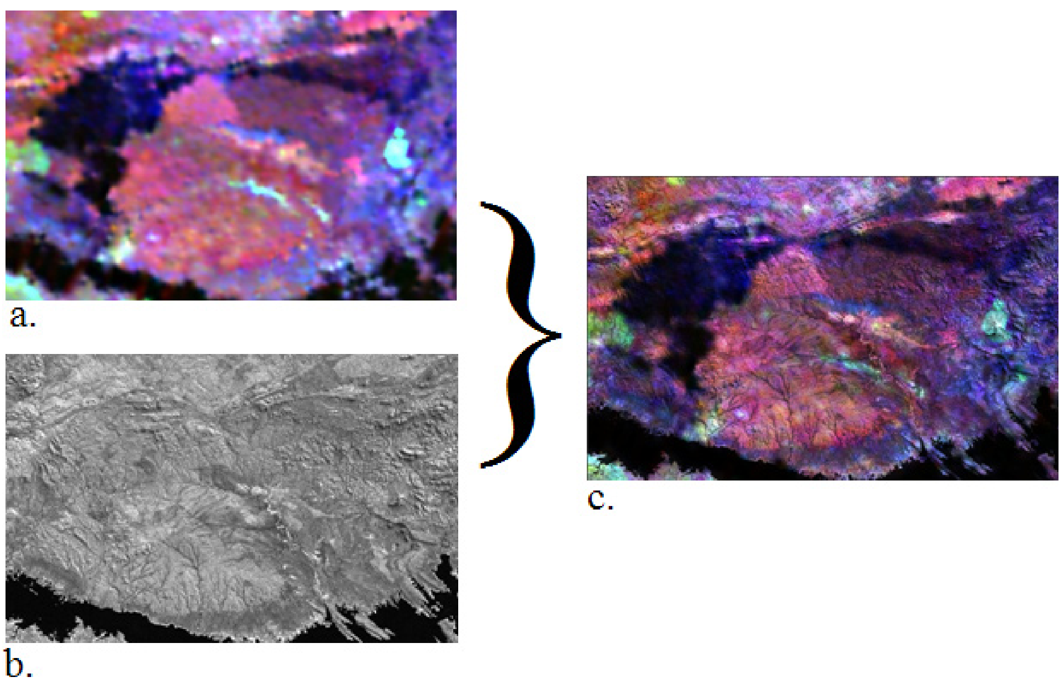


Figure 2. a. Ternary gamma ray spectrometer color composite image K-Th-U (RGB) – imagery supplied by E. Schetselaar – ITC) b. Landsat TM 7 c. Integrated image using Saturation-Value-Modulation (SVM) method.

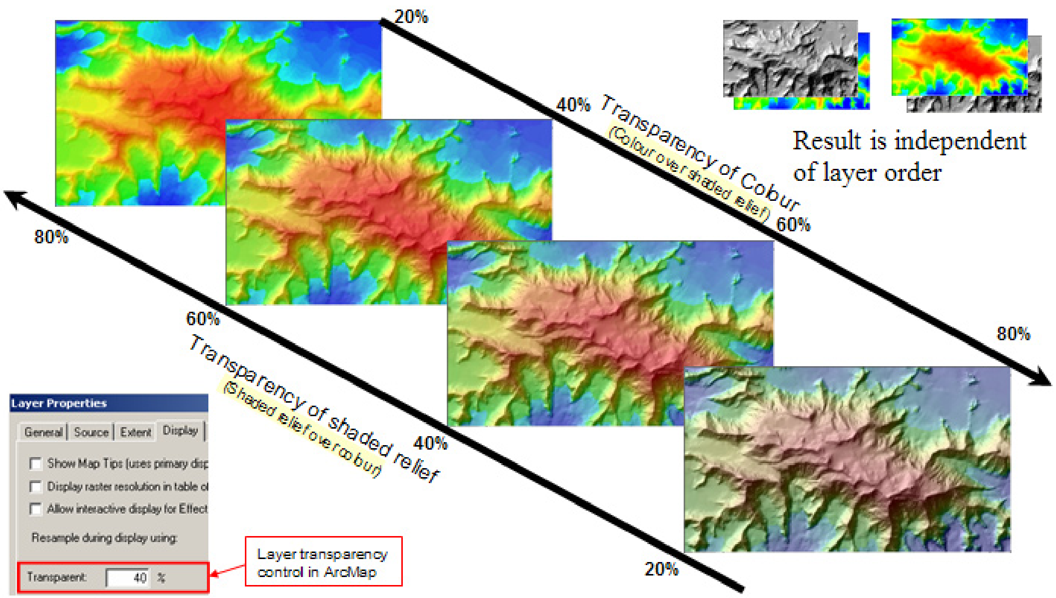


Figure 3. Layer transparency feature of ArcGIS.

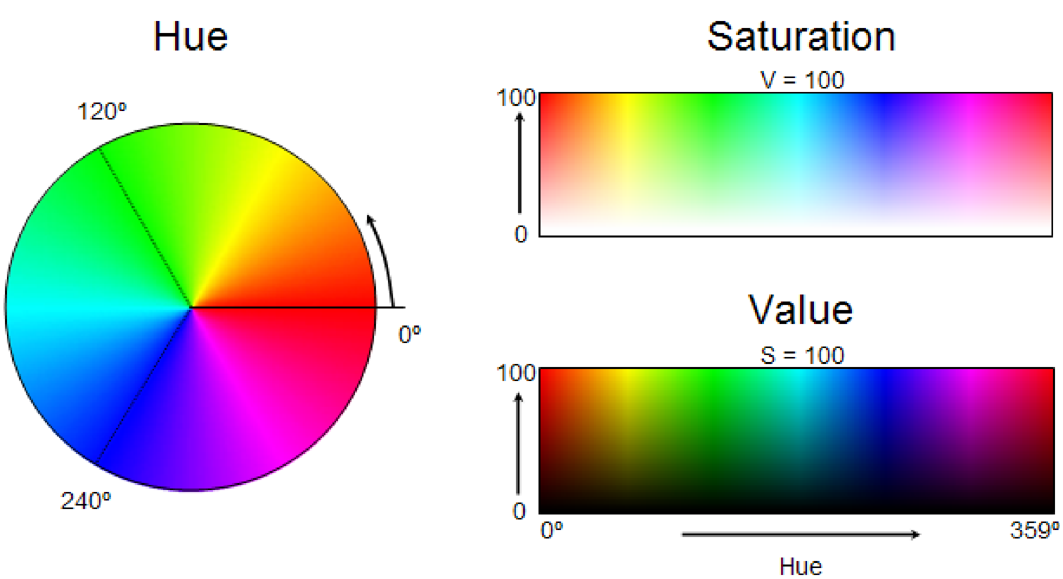


Figure 4. Hue-Saturation-Value color model.

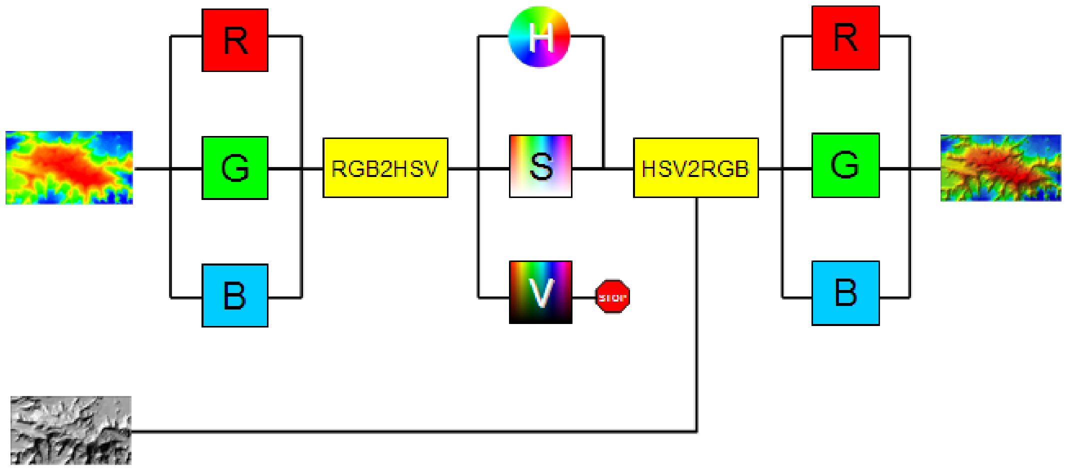


Figure 5. Traditional Value Replacement Method of image integration.

is, often the difference between lighter and darker colors in an image is higher and lower numbers representing the value component of the colors. If the value components are modulated or replaced, then the difference between the colors will be changed or eliminated. Figure 6 shows how replacing the value component of a dark and lighter green results in an image where the dark and lighter greens cannot be differentiated. In this example, the only difference between the two greens in the original color image is in the value component. A second problem with this technique is that the original colors can be corrupted when the value component is replaced (Harris et al., 1990, 1994). For example, yellow can appear as dirty green in the integrated image, and red can appear brown. Another problem with this method is that, if the saturation of the colors is low, then replacing the value component results in an image where the saturation of the color is further reduced (Figure 7).

Another traditional method of image integration involves pixel-by-pixel multiplication of the RGB components by the grayscale values scaled between 0 and 1 (Figure 8). The main problem with this method is that the scaled values of nearly all pixels in a grayscale image are less than one, so the colors in the resulting integrated image are darker than the original image. This scaled values problem can also corrupt the apparent hue of the color. For example, a yellow might appear to be some kind of green (Figure 9).

Both value replacement and RGB modulation methods offer very little flexibility on how the integration is performed, and there are few or no parameters that can be used to control the result of the calculations.

Unlike layer transparency, the SVM method is not interactive and involves pixel-by-pixel computations similar to those associated with the value replacement and RGB modulation methods. However, the SVM method provides more flexibility on the integration process which results in integrated images that are superior to those produced by traditional methods.

OVERVIEW OF THE SATURATION-VALUE-MODULATION (SVM) METHOD

The saturation and value color components of an illuminated object change with the angle of incidence. For example, Figure 10 shows a cylinder illuminated from the right side. The colors of the parts facing the source of illumination have a lower saturation and a higher value component, whereas the parts facing away from the source of illumination have a higher saturation and lower value. In the area around the "cutoff" line (Figure 10), the saturation and value components will be that of the natural color of the object.

In Figure 11, the "shade value" (x-axis) is zero on those surfaces that face away from the source of illumination; the highest values (e.g., 255 or 100) will be assigned to surfaces that face toward the source of illumination. Multiplier curves that range between 0 and 1 can be used to modulate the saturation and value color components depending on the shade value. The value of a pixel in the grey-scale image (Shade value) defines a vertical line that intersects the saturation and value multiplier curves. The points of intersection of this vertical line and multiplier curves are the saturation and value multipliers ranging from 0 to 1 (Figure 11). The saturation and value components of the color image at the same pixel location are multiplied by their respective multipliers. The resulting modulated saturation and value components are integrated with the original hue component to create the SVM image in HSV coordinate space. The final step is a transformation of the hue and modulated saturation and value components to RGB coordinates for display purposes.

Figure 12 presents a schematic of the SVM method. Not shown is the transformation of the color image from

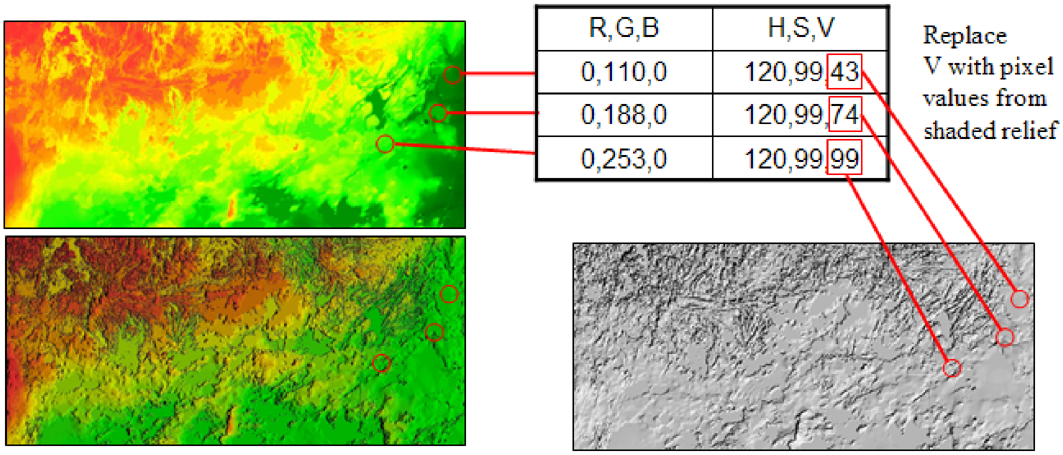


Figure 6. Loss and corruption of color with value replacement method. The three circled areas have three shades of green differentiated only by the value component as shown in the table. Replacing the value component of these green areas with values in the shaded relief image results in a loss of the shades of green. Replacing the value component in yellow and red areas results in colors that appear "dirty".

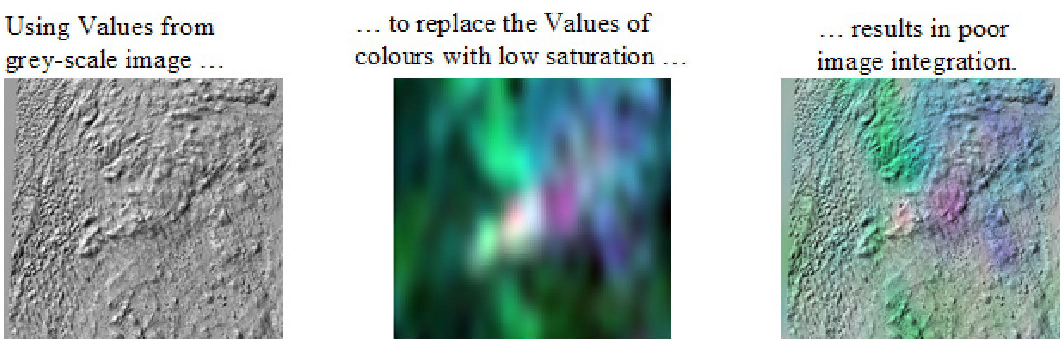


Figure 7. Low saturation colors become lower with value replacement method.

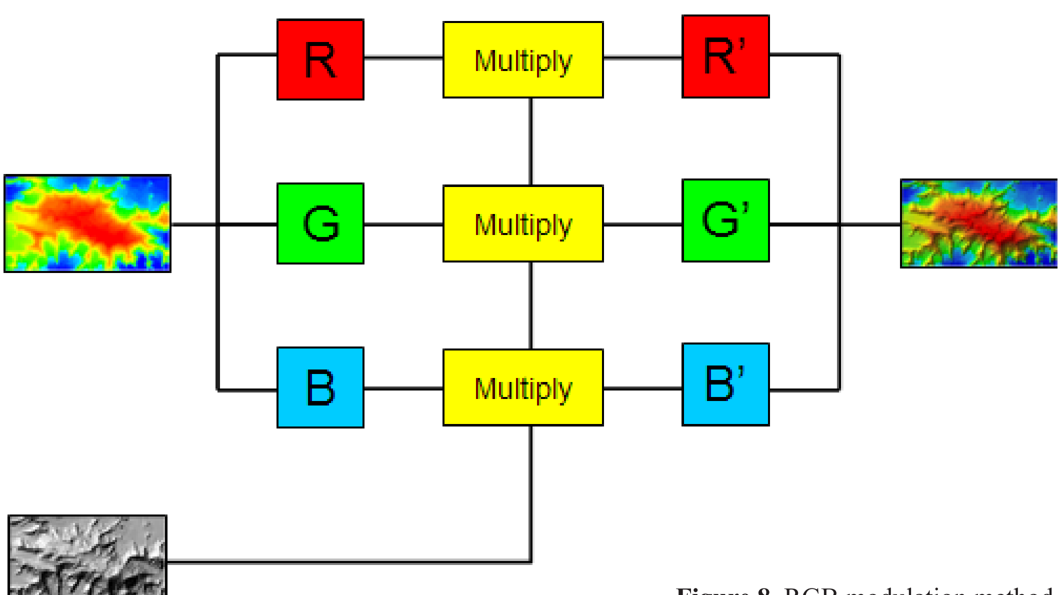


Figure 8. RGB modulation method of image integration.

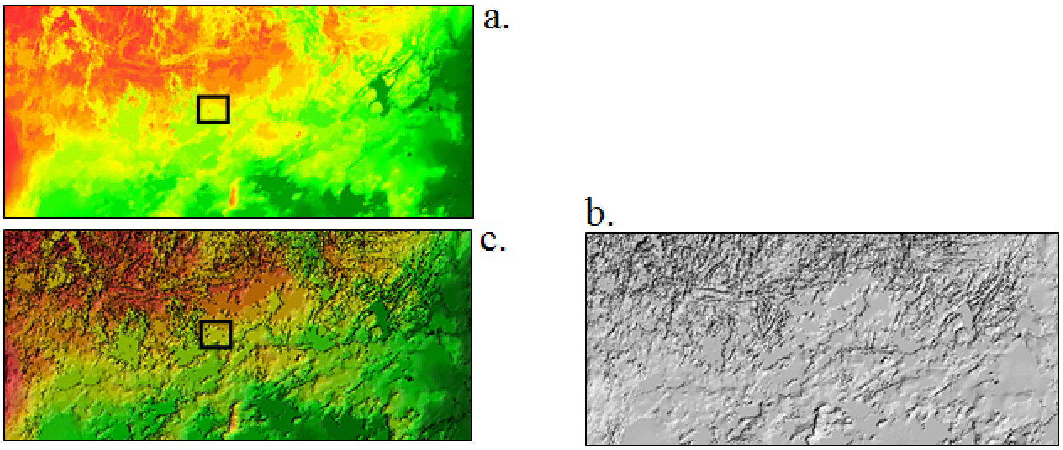


Figure 9. a. Color DEM image of northern Manitoba. b. Shaded relief of northern Manitoba DEM. c. Integrated image using RGB modulation method. The box highlights an area where light green, yellow, and light red pixels have been transformed to darker colors since all pixels have been multiplied by a value from the shaded relief image of less than 1. In fact, virtually all pixels will be multiplied by multipliers less than 1 which creates a darker overall integrated image product.

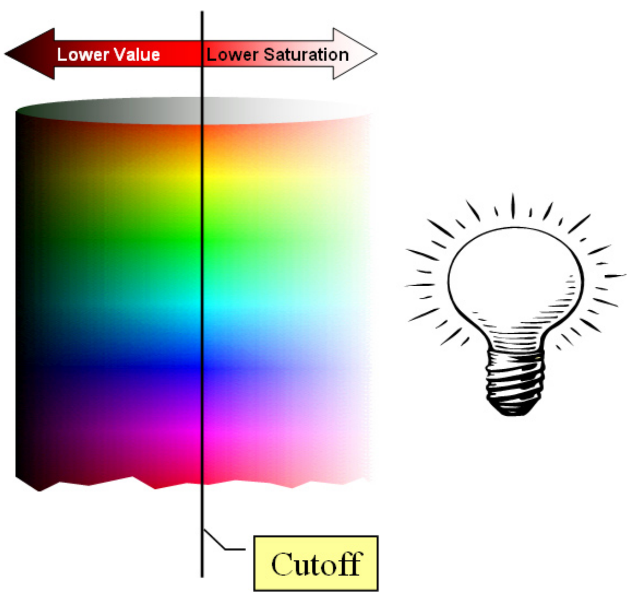


Figure 10. A rainbow colored cylinder illustrates the Saturation-Value-Modulation (SVM) Concept. For surfaces facing the illumination source, the colors have full value but lower saturation. For surfaces facing away, the colors have full saturation but lower value. For surfaces at the "cutoff", the color has full saturation and value.

RGB to HSV components. The schematic shows how the saturation (S) and value (V) components are modulated by multipliers that are determined by the pixel value in the grayscale image. The modulated saturation (Sm) and value (Vm) are used with the original hue component and transformed to an RGB composite image file that can be displayed in remote sensing software or a GIS. The multiplier curves and the parameters that define their shape are key elements of the SVM method.

SVM Parameters

There are five SVM parameters necessary to define the shape of the saturation and value multiplier curves. Together, they provide the ability to control various characteristics of the resulting integrated image.

Grayscale Value Cutoff (CutOff)

In cases where the pixel values in a grayscale image are lower than the "grayscale value cutoff" (see Figure 11), the value component (V) of the color image will be modulated, and the saturation component will be equal to the saturation in the original color image (i.e. saturation multiplier equals one). For grayscale pixel values greater

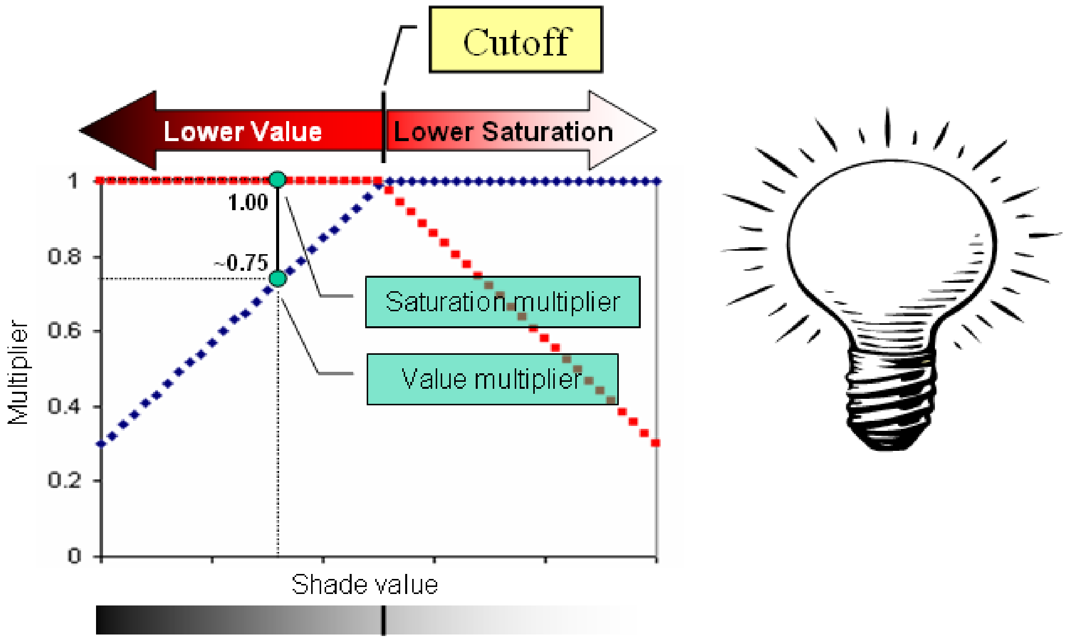


Figure 11. SVM saturation and value multiplier curves are used to model the lower saturations for surfaces that face towards an illumination source and lower color values for surfaces that face away from an illumination source. Surfaces that neither face towards nor away from the illumination source (e.g. horizontal surfaces in a digital elevation model) will have minimal or no change to their original color values.

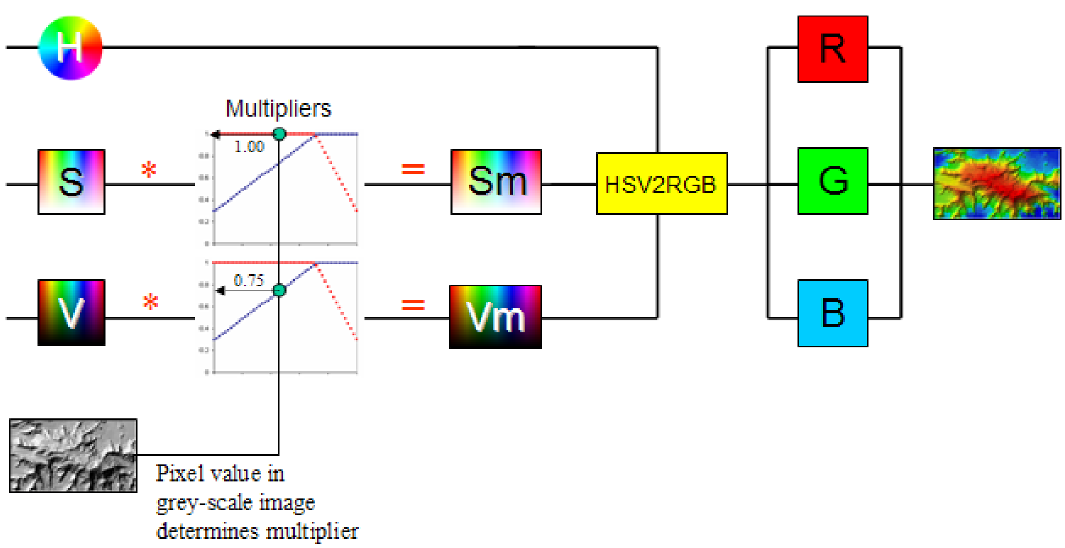


Figure 12. Schematic of the SVM method.

than the cutoff, the saturation component (S) will be modulated, and the value component will equal the value in the original color image (i.e., multiplier equals one). At the cutoff value, the value and saturation multipliers are equal to one, so the color in the integrated image will be the same as the original color image at the cutoff value.

A general SVM rule of thumb is to maximize the number of pixels in the integrated image that have the same color as the original color image. This means that a cutoff value that maximizes the number of pixels with saturation and value multipliers of 1 should be selected. In most cases, this cutoff is represented by the peak in the grayscale image histogram.

For a typical shaded relief digital elevation model, the peak in the histogram coincides with pixels that represent horizontal surfaces. For shaded relief images, the cutoff can therefore be computed from the following equation:

Cutoff = 255 * sin(A)

This assumes a range of values in the grayscale image is 255 and A is the illumination source altitude in degrees (0-90). For example, the peak of histogram of a shaded relief DEM with an illumination angle of 45 degrees will be approximately 180 (Figure 13).

Figure 14 shows the impact of changing the cutoff on the integrated image. Lowering the cutoff below the optimal value of 180—the peak in the grayscale histogram—results in an image with lower overall saturation (i.e., washed out colors) than the original color image. Increasing the cutoff value above the cutoff results in an image with lower overall value (i.e., darker).

Minimum Value Multiplier (Vmin)

The minimum value multiplier, which can vary between 0 and 1, determines how dark the pixels will be in areas where the grayscale pixel values are low. For example, if Vmin is 0, then for grayscale pixel values of 0 the value component of the color image will be multiplied by 0. Any color with a value component of 0 is black. This means that color pixels that have the same location as grayscale pixel values of 0 will be black in the integrated image. As Vmin increases, these same pixels will become brighter. A Vmin value of 1 will result in no value modulation. In this case, the pixels in the integrated image that have the same location as grayscale pixels with values lower than the cutoff will be the same as the original color image.

Figure 15 shows the impact of changing the value of Vmin. Note how the shadows become brighter as Vmin increases. Values greater than 0 and less than 0.4 are generally recommended.

Value Multiplier Exponent (Vexp)

The value multiplier exponent will increase or decrease the number of pixels that will have their value component multiplied by a value close to Vmin. Higher Vexp values mean that the multiplier will rise slowly from Vmin. Figure 16 shows how increases in Vexp increase the proportion of "dark pixels" in the integrated image. Values of 1 or less generally provide good results.

Minimum Saturation Multiplier (Smin)

The minimum saturation multiplier, which varies between 0 and 1, determines how much color there will be for pixels where the grayscale values are high. For example, if Smin is 0, then for grayscale pixel values of 255 (the maximum in the image), the saturation multiplier will be 0 and the saturation of the color in the integrated image will be 0. In this case the color in the integrated image will have "no" color and will typically be white or light gray. As Smin increases, these same pixels will have higher saturation and more color. If Smin is 1, then there will be no saturation modulation and the pixels above the cutoff will appear to be "flat". Figure 17 shows the impact of increasing Smin from 0 to 0.6. Smin values between 0 and .4 are recommended.

Saturation Multiplier Exponent (Sexp)

The saturation multiplier exponent will increase or decrease the number of pixels that will have their saturation component multiplied by a value close to Smin. Higher Sexp values mean that more pixels will be multiplied by a multiplier close to Smin. Figure 18 shows how higher values of Sexp decrease the proportion of "washed out pixels" in the integrated image. Generally values between 1 and 3 provide good results.

ARCGIS IMPLEMENTATION OF SVM

The SVM method was implemented as a Visual Basic for Applications (VBA) application in ESRI's ArcMap application. It works with ArcGIS (ArcView, ArcEditor, or ArcInfo) and does not require any special ESRI extensions (e.g., Spatial Analyst).

The VBA implementation allows the user to set each of the five SVM parameters described above. The input images can be one of the following:

- 8- or 16-bit image with an associated CLR file
- 8- bit TIF image (colors embedded in file)
- 32-bit TIF image (color defined by RGB channels)
- 3 separate 8-bit RGB images

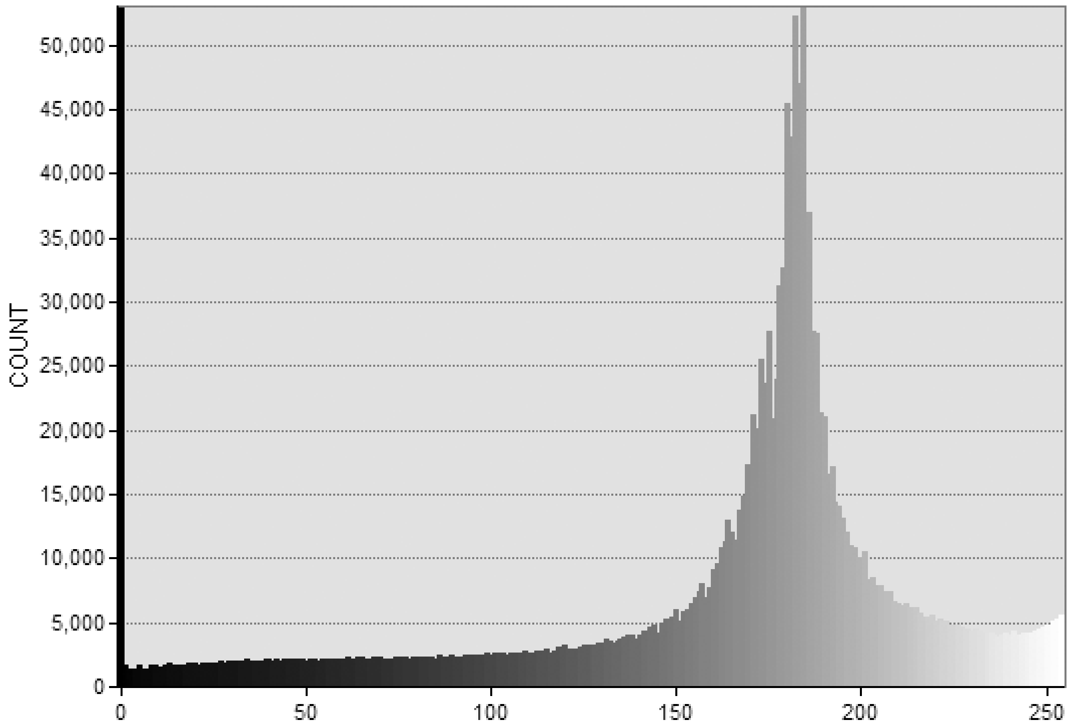


Figure 13. Histogram of pixels values for shaded relief of Mt. Logan (Figure 1). Peak is approximately 255 * sin(45) = 180 = recommended SVM cutoff value.

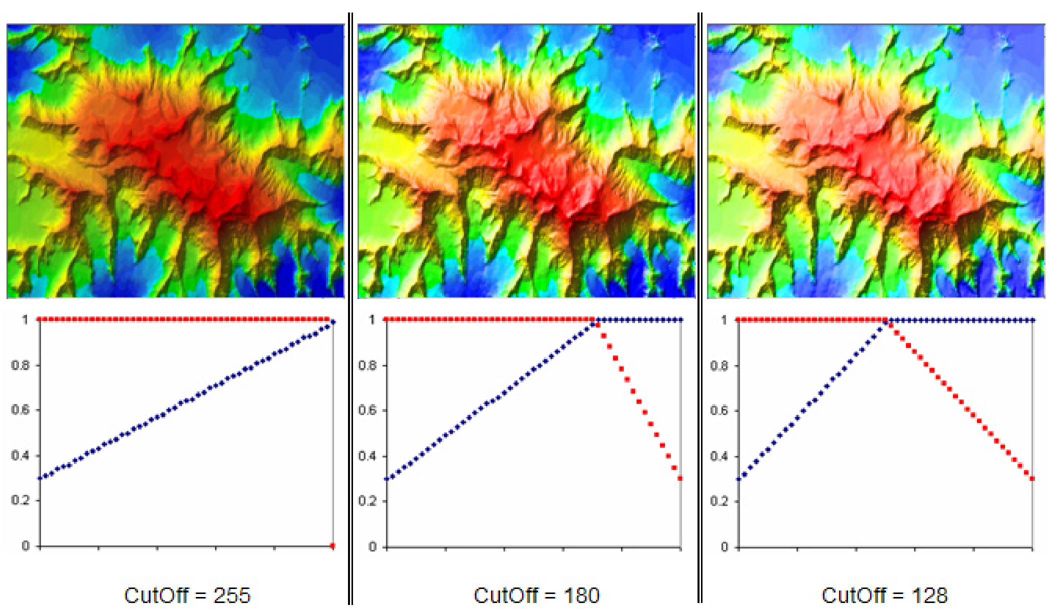


Figure 14. SVM Parameters – Grayscale value cutoff (CutOff).

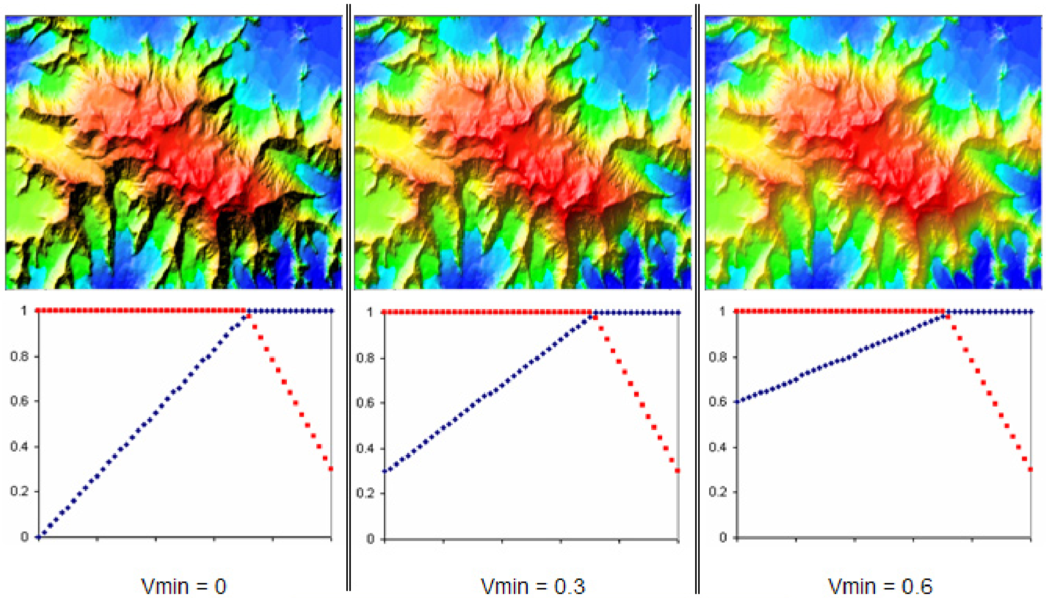


Figure 15. SVM Parameters – Minimum value multiplier (Vmin).

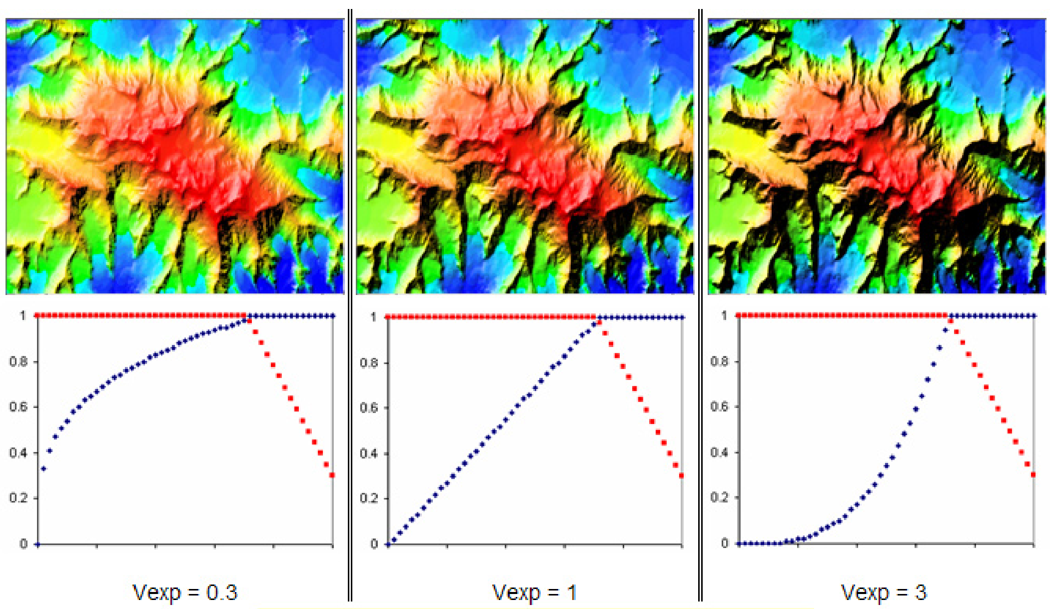


Figure 16. SVM Parameters – Value multiplier exponent (Vexp).

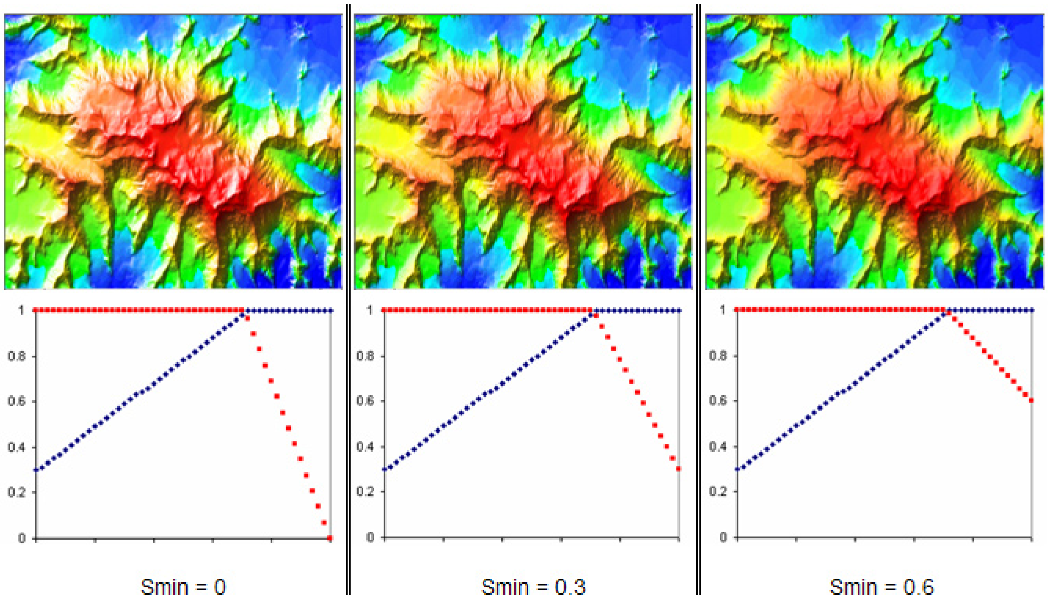


Figure 17. SVM Parameters – Minimum saturation multiplier (Smin).

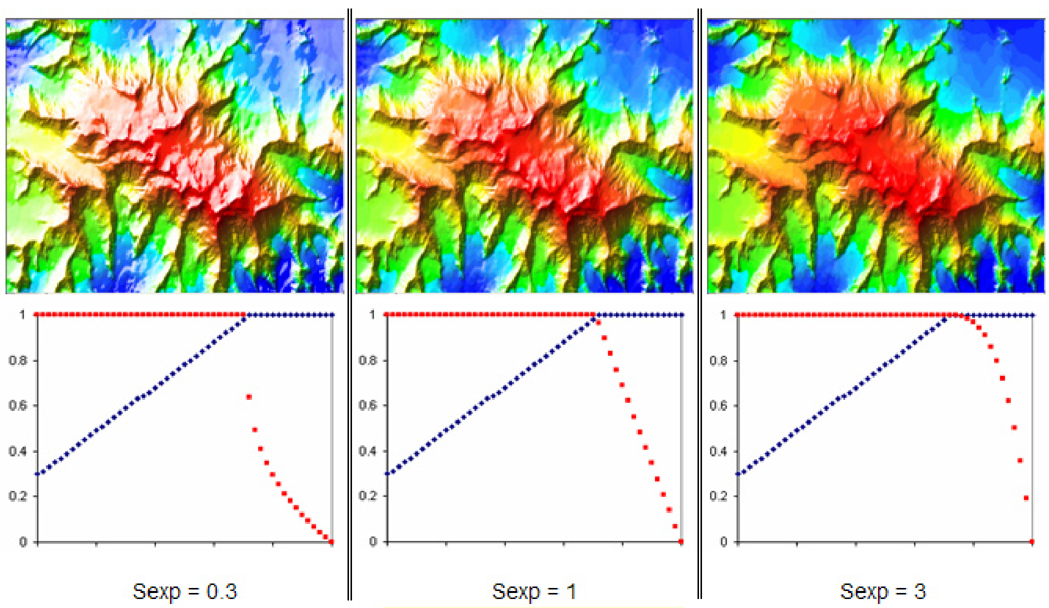


Figure 18. SVM Parameters – Saturation multiplier exponent (Sexp).

The CLR file contains space delimited values for pixel value and RGB components. An example record from a CLR file might be:

18 244 64 120

This means that pixel values of 18 in the color image have RGB coordinates of 244, 64, and 120 respectively.

The output from SVM is a 3-band RGB Band Interleaved by Line (BIL) image which is easily imported or used directly by remote sensing software. It can also be easily exported to TIF or ESRI Grid format in ArcGIS.

The ArcMap document (MXD) that contains the VBA code can be downloaded from the SVM FTP site along with sample data (Viljoen, 2006).

APPLICATION OF SVM IN GEOSCIENCE

The SVM method has broad application to geoscience studies that require integration of a color and grayscale image. Many geological mapping applications, for example, require interpretation of various types of remotely sensed and geophysical data. The integration of these data types often provides images that offer a unique perspective of the Earth's surface, which enables the interpretation of many geological features that, without integration, would have been impossible to make. Furthermore, the relationships often evident in the resulting integrated imagery offer a unique interpretation tool. The following three examples highlight the value of integrating different geoscience data and demonstrate the advantages of the SVM method over traditional integration methods.

Integrating Geological Map Units with Shaded Relief Aeromagnetics

The magnetic characteristics of rocks at, and below, the Earth's surface often reflect mappable variations in lithologies. The magnetic characteristics of rocks are measured with aeromagnetic sensors, and these measurements are often processed into colorful images that represent total field, vertical gradient, and other derivative products. Integration of colored geological units with shaded relief versions of total field aeromagnetics can provide an image that is extremely useful for geological mapping.

Geological units are usually represented by vector polygons in a GIS and, given that SVM is entirely a raster-based method, these vector polygons must be rasterized and have the same projection, pixel resolution, and map extent as the total field shaded relief image. Detailed step-by-step instructions on how to use the ArcGIS version of SVM for this kind of integration are available for download (Grant and others, 2006).

Figure 19 shows the result of using SVM to integrate rasterized geological units with shaded relief total field aeromagnetics. As can be seen, rock units, after SVM integration, are characterized by different magnetic signatures on the integrated image. This image

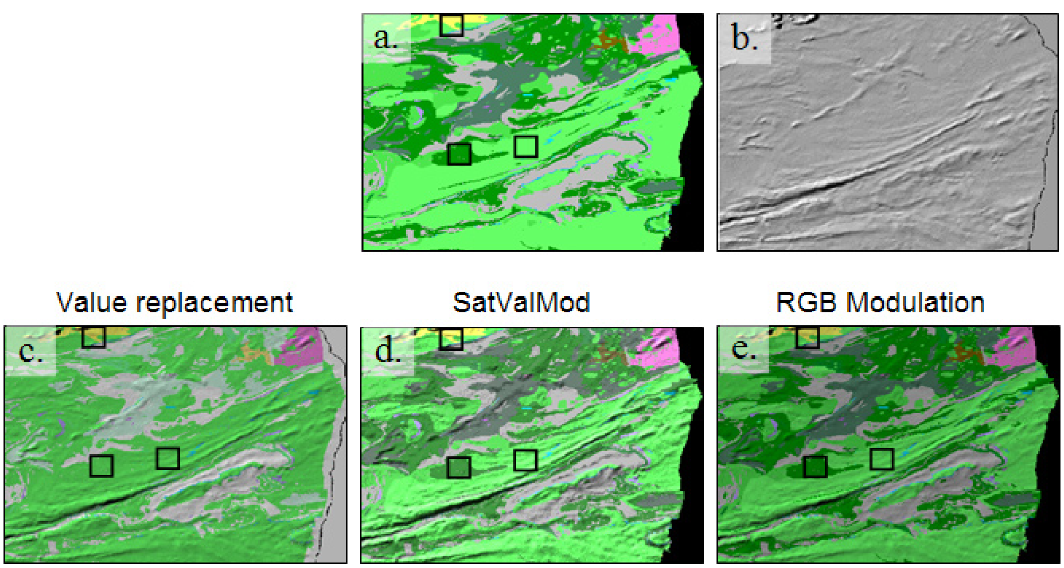


Figure 19. Comparing image integration methods – a. Rasterized geological map units. b. Shaded relief aeromagnetic image for the same area as the geology map. c. Integrated image using the value replacement image integration method. d. Integrated image using the SVM method. e. Integrated image using RGB modulation method.

can greatly assist mapping endeavours, as the rock units can be modified based on variations in their magnetic signature. With respect to the integration methods, note how the integrated image produced by the SVM method retains the original colors of the geology map compared to the other two traditional methods. The small squares highlight areas where color corruption and loss are evident in the integrated image produced by the other methods.

Integrating a Ternary U-Th-K RGB Gamma Ray Spectrometer Image with a Shaded Relief Digital Elevation Model

Integration of gamma ray spectrometer data, which measures the emission of the elements U, Th and K from the Earth's surface with a shaded DEM, can also provide a very useful image for geological mapping. Variations in the above radioelements often reflect different rock units as well as areas of potential mineralization. However, since the gamma ray data comprise three channels (U, K, Th) that are often correlated, the color variations in ternary images are often low. The integration of ternary imagery presents a challenge to all methods of image integration. The Value replacement method, for example, often results in almost complete color loss as shown in Figure 20. As with the previous example, the SVM method results in an integrated image that retains the original color to a much greater degree than the other methods.

Pan Sharpening Landsat Multi-spectral With A Panchromatic Image

Pan sharpening is the term given to an image processing technique that uses higher resolution (smaller pixels) grayscale imagery to improve the visualization of lower resolution (large pixels) color or color composite images. Many satellite and airborne sensor systems have multi-spectral channels and panchromatic channels. Landsat 7, for example, provides 6 channels of 30 meter pixel resolution for portions of the visible, near infrared and short-wave infrared of the electromagnetic spectrum. It also contains a panchromatic channel that covers the entire visible part of the electromagnetic spectrum with 15 meter pixel resolution. Integrating the higher resolution panchromatic channel with a RGB color composite of the lower resolution visible channels results in a sharper color composite image as shown in Figure 21. The boxes in the images are areas that highlight the differences between these image integration methods. If the accuracy of the original colors is important, than SVM is the preferred method. If the original colors are not important, then value replacement is another pan-sharpening option.

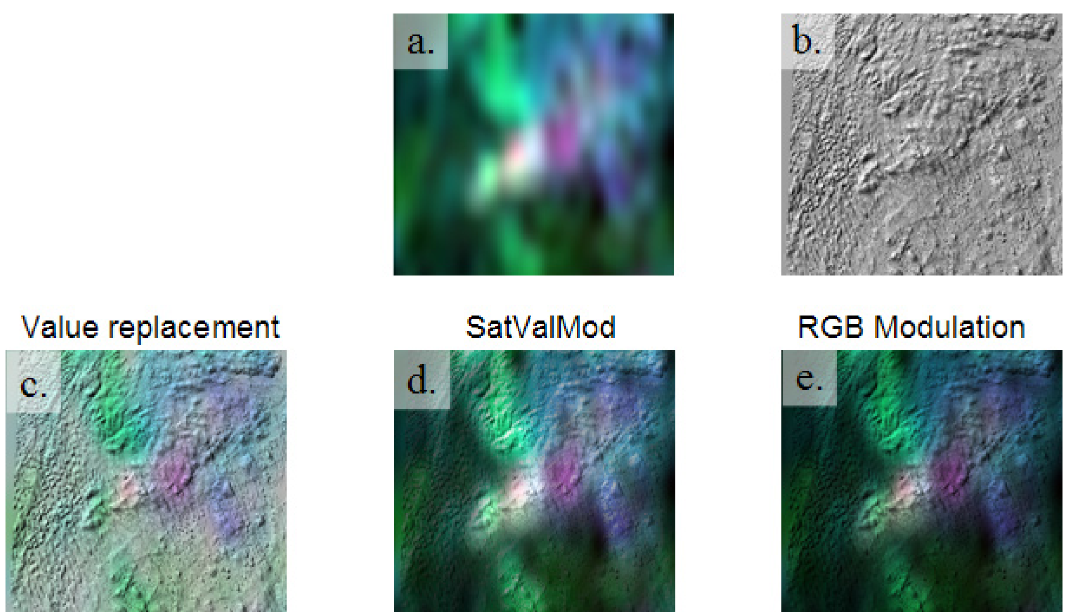


Figure 20. Comparing image integration methods – a. Ternary gamma ray spectrometer color composite image K-Th-U (RGB) composite image. b. A shaded relief digital elevation model of the same area as the K-Th-U image. c. Integrated image using the value replacement image integration method. d. Integrated image using the SVM method. e. Integrated image using RGB modulation method.

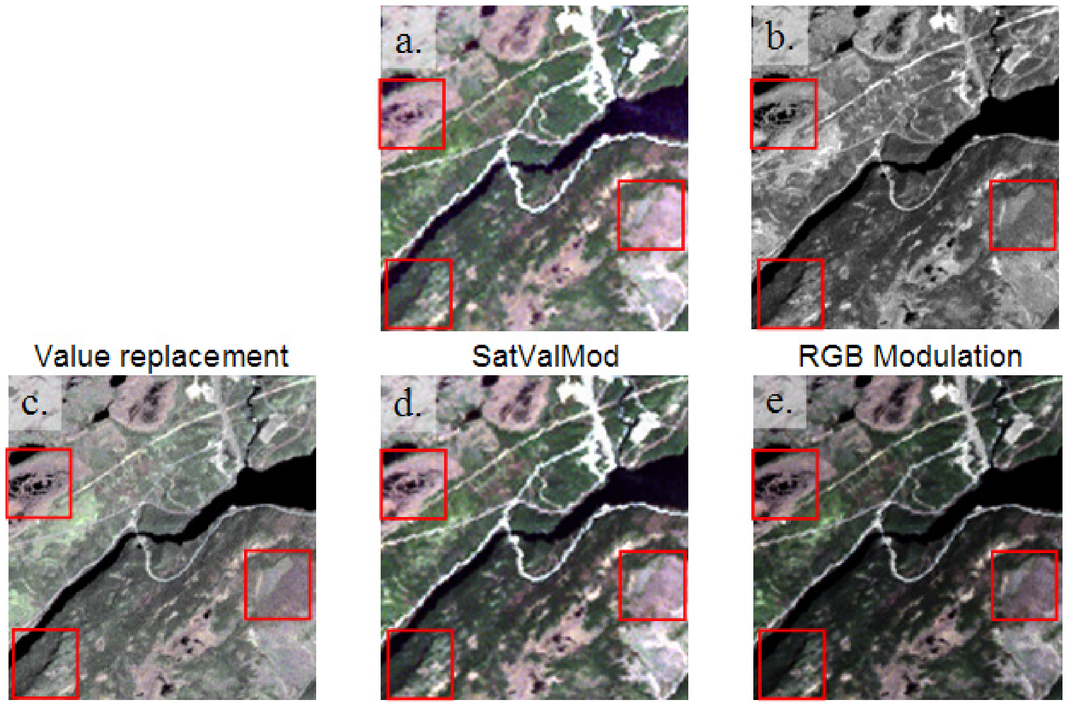


Figure 21. Comparing image integration methods. a. Landsat 7 RGB color composite image of Thematic Mapper (TM) bands 3, 2, and 1 respectively (30 meter pixels). b. Landsat TM Band 8 (15 meter pixels) c. Pan sharpened image using the value replacement image integration method. d. Pan sharpened image using the SVM method. e. Pan sharpened image using RGB modulation method.

CONCLUSIONS

The Saturation-Value-Modulation (SatValMod or SVM) method is based on the real-world concept of darker colors for surfaces in shadow (lower value component) and color loss (lower saturation component) on illuminated surfaces. This real-world concept is implemented in SVM as a pair of multiplier curves that modulate the saturation and value components of colors in the color image. These modulated saturation and value components are combined with the original hues in the color image and transformed to red-green-blue components for display.

The SVM method of integrating color and grayscale imagery provides superior results over many other integration methods because there is no compromise between color and shading, as is the case for layer transparency. In addition, there is no distortion of colors, as can result from value replacement, RGB modulation, and other methods. Unlike the other integration methods, which provide little or no control over the integration process, SVM uses five different parameters that provide a great deal of control over the characteristics of the resulting integrated image.

The use of color is important in conveying geoscience information such as geological units, geophysical properties, radiometric characteristics, and many others. Visualizing the relationships between these data is made possible through image integration techniques. The SVM method is a superior algorithm for integrating color and grayscale imagery, which results in integrated images that preserve the original color and grayscale characteristics of the input imagery. SVM is freely available from the SVM FTP site listed in Viljoen (2006).

REFERENCES

Harris, J.R., Murray, R., and Hirose, T., 1990, IHS Transform for the Integration of Radar Imagery with Other Remotely Sensed Data: Journal of Photogrammetric Engineering and Remote Sensing, Vol. 56, No. 12, p. 1631-1641.

Grant, G., Fraser P., and Viljoen, D., 2006, How to SVM: accessed at ftp://nrd:imanrd2@ftp.gis.nrcan.gc.ca/viljoen/downloads/satvalmod/HowToSVM.pdf.

Harris, J.R., Bowie, C., Rencz, A.N., and Graham, D., 1994, Computer Enhancement Techniques for the Integration of Remotely Sensed, Geophysical and Thematic Data for the Geosciences: Canadian Journal of Remote Sensing, Vol. 20, No. 3, p. 210-221.

Viljoen, D., 2006, Saturation-Value-Modulation FTP site: accessed at ftp://nrd:imanrd2@ftp.gis.nrcan.gc.ca/viljoen/downloads/satvalmod/.

Unpublished Geologic Evidence and Other Databases of the Kentucky Digital Mapping Program

By Gerald A. Weisenfluh and Douglas C. Curl

Kentucky Geological Survey
228 Mining and Mineral Resources Bldg.
University of Kentucky
Lexington, KY 40506-0107
Telephone: (859) 257-5500
Fax: (859) 257-1147
e-mail: {jerryw, doug}@uky.edu

INTRODUCTION

Between 1996 and 2005 the Kentucky Digital Mapping Program digitized 707 USGS Geologic Quadrangle maps produced by the joint USGS-KGS geologic mapping program that took place from 1960 to 1978. One of the products of this effort is a seamless 1:24,000-scale spatial database of geologic information. The database includes map features and all the explanatory information found on the collar of the printed USGS maps. Since 2001, the Kentucky Geological Survey has been developing Web services to integrate the digital geologic maps with other geoscience databases (Weisenfluh, and others, 2005).

The seamless, 1:24,000-scale database is essentially complete and available as an ESRI ArcIMS map service at the Kentucky Geologic Map Information site (http://kgsmap.uky.edu/website/KGSGeology/viewer.asp). Some supporting data are still being added to the site. Here, we discuss database design elements that were found to be particularly useful for implementing the Web project and current KGS database activities that will extend the usefulness of the Internet mapping site.

MAP EXTENT DATABASE

The first database that was built to support the geologic map system was a catalog of KGS and USGS reports and maps. Map coordinate extents were assigned to publications to facilitate geographic searches for information of this kind, where a single coordinate value does not adequately represent the area to which the information applies. At the onset, most publications were associated with quadrangles or counties, but subsequently, additional area types such as state parks and drainage basins were added. A database of map extents was constructed that contains a geographic area type and name along with the minimum and maximum coordinates for the enclosing rectangular area.

At about the same time, Internet map developers in other Kentucky state government agencies were creating new Internet map services for a variety of information. Among the early drawbacks to these maps were the slow redraw speeds and the difficulty of finding a specific area on a statewide map. Most users utilized repeated zoom and pan methods to find an area, and so finding an area of interest was taxing for both the user and the computer servers. KGS used the map extents database to develop a search and zoom function that could be used by any Kentucky Internet map developer to simplify this process. This service, called the KGSGeoPortal, functions much like the Geographic Names Information System with the added ability to link to most Kentucky-based Internet maps from a single base map. Moreover, the initial view provided by the GeoPortal is typically closer to the user's area of interest because the database stores the full extent of the feature, not just its central value. The GeoPortal is found at <http://kgsmap.uky.edu/website/KGSGeoPortal/KGSGeoPortal.asp>.

The database was enhanced to help resolve similar geographic names. For example, many stream names are duplicated in different parts of the state. Additional attributes were assigned to identify the beginning and ending county name, the drainage basin name, and the name of the stream of the next higher order so that users could identify the correct stream.

All of the initial map extents for this database were for standard geographic features that were readily available in GIS format. The extents were calculated using a simple Avenue function in ArcView or with AML programs in Arc for more complex requirements. More recently, the need to store informally-defined areas has increased, but it was more difficult to calculate these areas. For example, many reports are published for study areas that do not conform to a standard quadrangle or other named geographic features. KGS undertook a project to help the Kentucky Transportation Cabinet catalog over 5,000 of its geotechnical reports for roadways and structures. Because each of these projects pertains to a small and unique area, an Internet base map service was developed with a function to create a custom map extent by dragging a rectangle across the area as viewed on a topographic map or aerial photograph (Figure 1). Most recent Transportation Cabinet project areas can be defined by the drillholes taken at the site, and therefore, another function was provided to upload the drillhole coordinates and plot them on the map. Users can then draw the project extent on the map or use the minimum and maximum hole coordinates to define the rectangle programmatically.

An unanticipated benefit of this work was discovered late in the project. The Transportation Cabinet has tens of thousands of historical drillholes, but only a few are referenced in a geographic coordinate system. Transportation engineers typically use a local survey system of route alignment station footages and offsets. Locating these holes on a map normally would be time prohibitive. Because the holes are associated with a project report and the reports have been assigned a map extent that, in most cases, is only hundreds of feet square, most of these data can now be placed in a geographic context that is suitable for comparison to geologic maps.

MAP DESIGN DATABASE

The KGS Geologic Map Information site was designed so that users could create a highly customized geologic map of a project area overlain with other related site information, such as oil and gas wells or sinkholes. Over 40 themes are available for geology, derivative classifications of map units, well and sample sites, economic features, hazards, and various base maps. Geologic maps can be customized by selecting individual layers or by choosing one of five predesigned layouts (Figure 2). The service also has a bookmark function that allows users to save a browser "favorite" for the current view that includes the map extent and the map layout. With a bookmark, users can return to the map exactly as designed or send it to a colleague. Bookmarks are stored as URL text strings, and saving in that string the necessary information about the visibility of 40 layers is not practical due to length considerations. Therefore, a database of map layouts (the "IMSLayers" database) was created so that a custom configuration could be stored in the URL as a simple variable number.

Columns in the IMSLayers database (Figure 3) designate the visibility of individual layers, while each

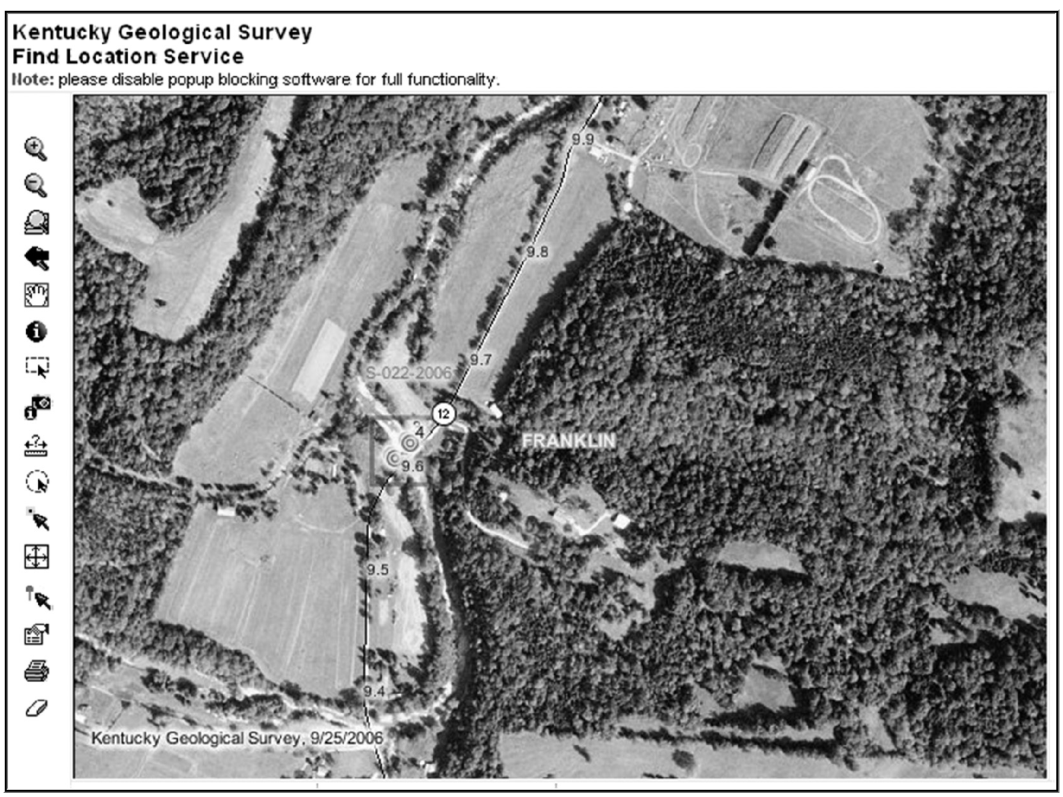


Figure 1. Creating non-standard map extents by interactively drawing a rectangle around plotted drill hole locations on an ArcIMS map service. The holes are plotted by means of a coordinate upload function.

row represents a unique combination of visible layers. To simplify map development, database column names for layers are equivalent to layer names used in the ArcIMS AXL file used for map rendering. Additional columns in the database record the layout ID number (an internal ID), a layout name for standard maps (this field is blank for map layouts defined by users), a counter for the number of users who have used that layout, and the name of the help file that is available for the layout on the map. For simplicity, the geologic map service has only one link to a help file for explanatory information. Because the content of the map can change significantly, the database can specify a different help file dynamically according to which layers are visible. For example, a karst help file is activated when the karst potential theme is turned on.

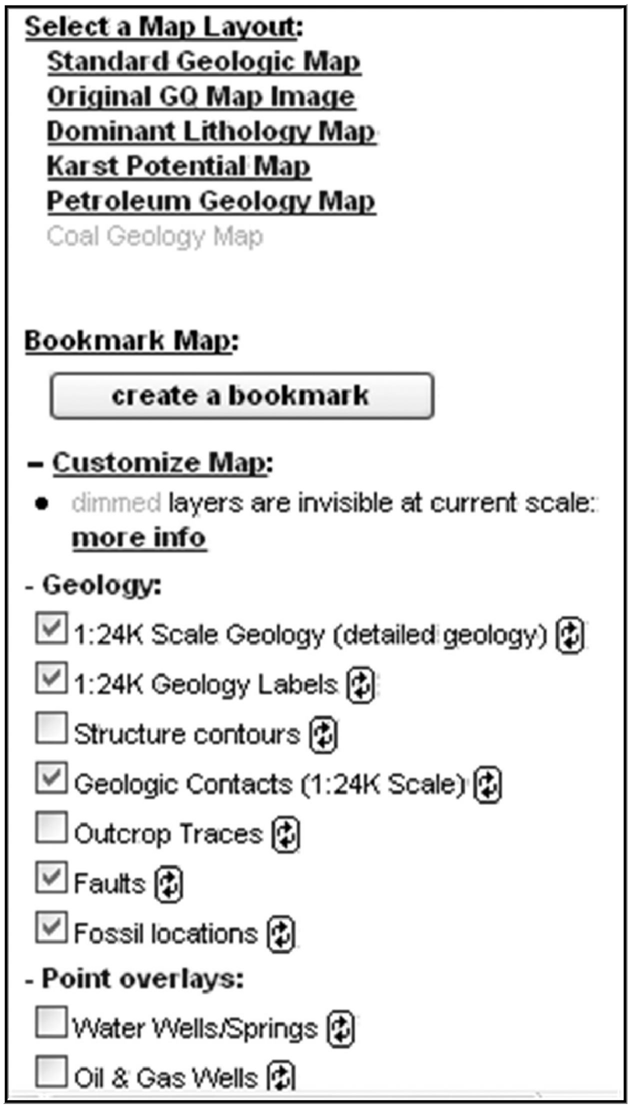


Figure 2. Pre-designed and customized map layouts on the KGS geologic map service.

When a user requests a bookmark, the system compares the current state of visible layers to the database. If the same layout exists, its ID number is appended to the URL along with the coordinate bounds of the view, and its count is incremented. If the layout does not exist, a record is appended to the table, and the new ID number is returned. When an Internet browser receives a URL with an embedded layout ID for this map site, the process is reversed. The ID number is used to obtain visible layers from the database, and the application draws the map in this initial state.

The IMSLayers database provides an easy method of storing customized map layouts that are useful to a diverse user base. Tabulation of layout counts provides information about frequency of use that can be used to identify "favorite" designs so that these can be provided as quick links.

DERIVATIVE GEOLOGIC MAPS

One of the objectives of the KGS Digital Mapping Project was to make geologic maps accessible to a wider user audience. In Kentucky, this includes many people without formal geologic training. One way to do this is to make derivative classifications of the geology that are conveyed in simpler terms appropriate to the target audience. Derivative maps were made using a variety of methods and are available on the map service as predefined layouts.

An oil and gas map was created by simple layer manipulation. This layout excludes most of the geologic themes, except faults and near-surface structure contours, then overlays oil and gas wells and fields. A dominant lithology map was created by reviewing lithology descriptions for each map unit, then assigning an appropriate

LayoutId	LayoutName	LayoutCount	HTMLFile	Ggeopoly	Gfault	Glabel	Gcontact	Gscontour	Gsmeasure	Gbed	Gfossil	Gxsect	Gcrop	Pwater	Poil
1	Standard Geologic Map	881	default_help.htm	-1	-1	-1	-1	0	0	-1	-1	0	0	0	0
2	Petroleum Map	372	default_help.htm	0	-1	0	0	-1	0	0	0	0	0	0	-1
42	Original GQ Map	273	default_help.htm	0	0	0	0	0	0	0	0	0	0	0	0
31	Karst Potential	277	karst_help.htm	0	-1	0	-1	0	0	0	0	0	0	0	0
3	Dominant Litho	192	litho_help.htm	0	-1	0	-1	0	0	0	0	0	0	0	0
33		0	default_help.htm	0	-1	0	0	-1	0	0	0	0	0	0	-1
34		2	default_help.htm	0	-1	0	0	-1	0	0	0	0	0	0	-1

Figure 3. Design of the IMSLayers database table. Layers assigned the value "-1" will be made visible.

term in a stratigraphic database. Combined units were treated similarly to heterogeneous stratigraphic units. Lithology terms were reduced to seven unconsolidated types and 23 rock terms, all of which were derived from the original geologic map legends. A karst potential map was developed using a multivariate analysis of factors that affect the dissolution of limestone units. This classification, developed by KGS geologists, is intended to show the likelihood that a unit may develop karst features. The four factors considered were percentage $CaCO_3$ in the carbonate portion of the unit, carbonate grain size, bedding thickness, and percent insoluble material. Each factor was ranked for all geologic units, and a combined score was tabulated. From this ranking, a five-level classification was devised, and adjustments were made based on the experience of KGS geologists. Examples of these map types are shown in Figure 4.

Two additional map classifications are in progress. The first is a map that depicts shale behavior by grouping units according to the amount of shale present and properties such as slake durability, expandable clays, and sulfide content. The second is an endangered species potential map that will highlight units that are associated with occurrences of endangered plant and animal species.

IMAGE LIBRARY

One of the most useful ways of communicating the character of geologic units is with photographic evidence. In fact, most geologists' offices are filled with hundreds, if not thousands, of photographs of outcrops, scenes, and specimens that have been accumulated during their careers. The challenge of making these images available to a general audience always relates to the time needed to catalog them—few working geologists have sufficient time. Recognizing this, KGS developed a digital image application for its staff that simplifies some of the tasks related to the processing and cataloging of photos. The application has Web-based data entry and search functions to provide ease of user access and efficient programming maintenance. The search function < http://kgsweb.uky.edu/geology/image_search.asp > is currently available for public access.

The initial step in cataloging a photograph is to upload the image to the Web server and, optionally, overprint credit text on the picture. No limits to resolution are imposed. Images are stored in their original resolution, and a thumbnail version of 100-pixel width is created. Those who wish to include a credit on the image can type

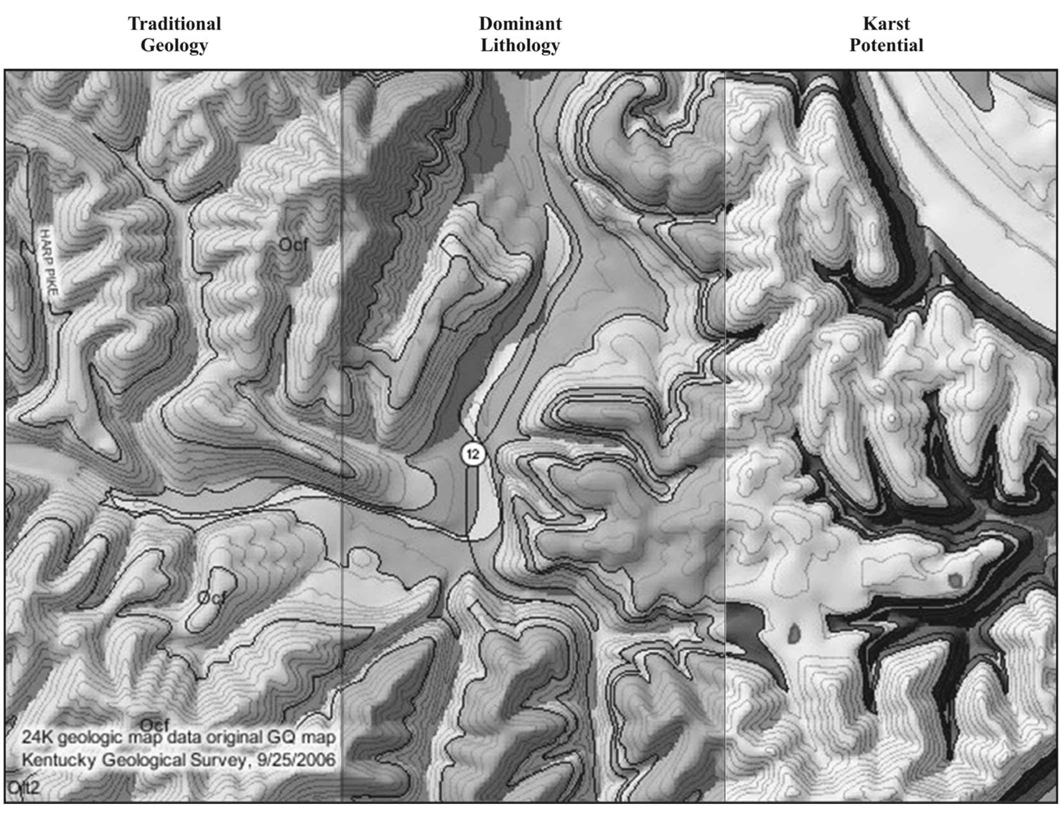


Figure 4. Examples of derivative classifications provided for Kentucky geologic maps.

the text, then specify the font size, color, and position (upper or lower left corner). A preview function is provided to verify that the selected options work well for the given image. Once the user is satisfied, the image is uploaded using the program AspUpload (Persits Software Inc., http://www.aspupload.com/index.html), and the credit text is appended during the operation. Some properties, such as pixel resolution and file size, are harvested from the image file, and a unique file name is assigned at the time of upload.

After the upload is complete, a form is presented for the geologist to describe the image. One of the most important characteristics to record is the location of the photograph. Locations can be specified with a variety of methods and levels of precision, ranging from "in a quadrangle" to a GPS coordinate value. Only images with a coordinate location can be posted on an Internet map; however, those with general locations can be searched from a map using the same methodology that is applied to publications (i.e., by assigning a mapextent value as described above).

The remainder of the cataloging process allows users to supply information about stratigraphic context, authorship, and image properties. It also allows users to provide brief and long descriptions of the photo, and assign keywords. Standard keywords are provided as checkboxes to simplify data entry and maximize the efficiency of the search function. Another keyword function is supplied to assign fossil names from a hierarchical taxonomic list.

The image search function (Figure 5) permits users to search by combinations of all the criteria discussed above. The initial results screen (Figure 6) contains a list of matching photos with a brief description, a thumbnail image, and links to a full description and a geologic map view of the location. Images can be downloaded in their original resolution. Users of the geologic map service can view photo locations on the map, and these symbols are linked to the detail description (download) page.

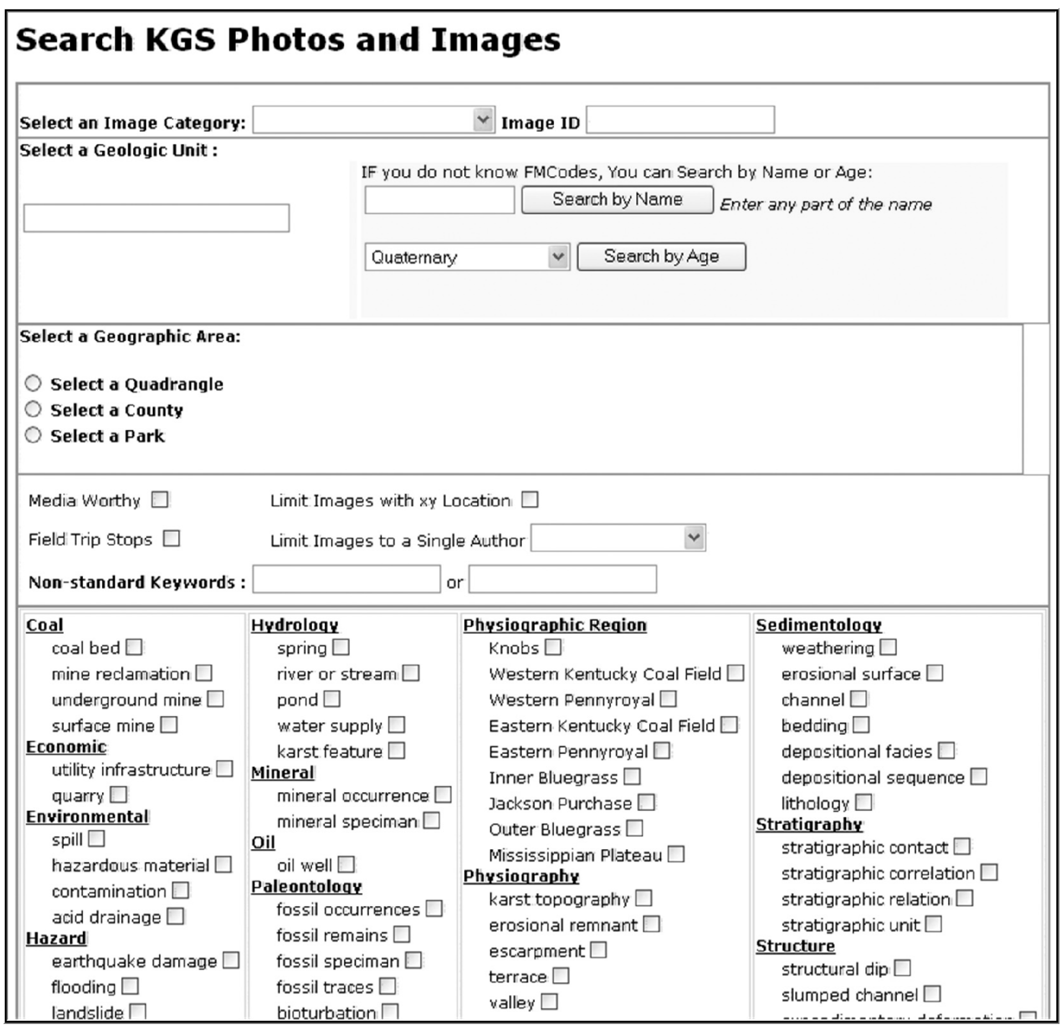


Figure 5. Part of the KGS photo search Web site.

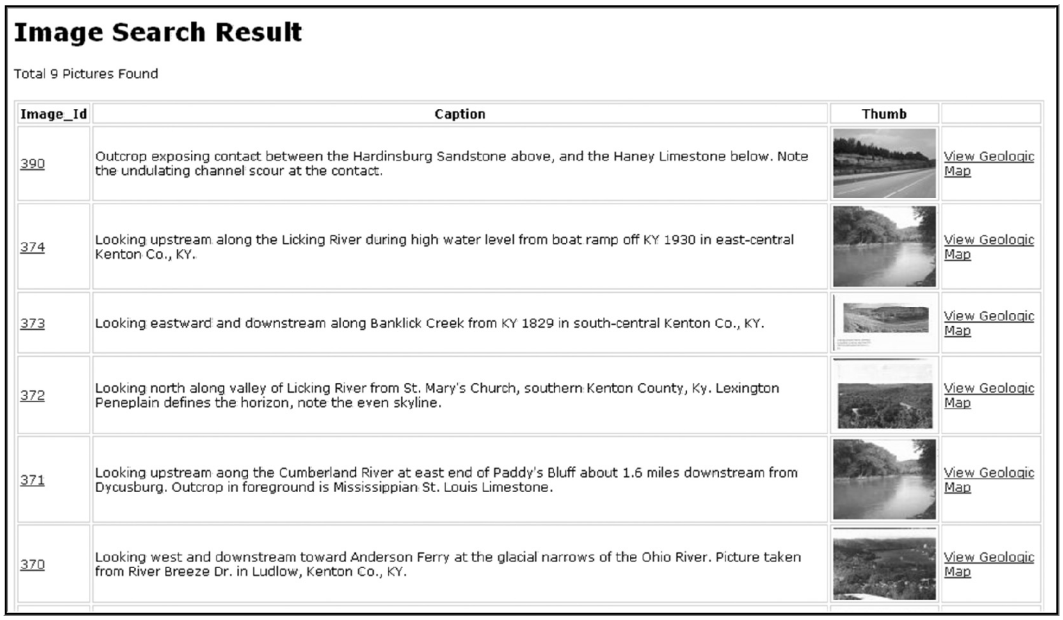


Figure 6. Photo search results page with links for detailed image information and a geologic map location.

GEOLOGIC ANECDOTES

The geologic map service provides information about the geology of Kentucky from published geologic maps, reports, and databases. However, some sources of information are unpublished, yet provide important context or explanation about the geology of an area. Unpublished evidence may include field notes from a mapping project, but often is represented as personal knowledge in anecdotic form. The Kentucky Survey sought a way of preserving this "institutional knowledge" and making it accessible to others. A simple form was developed to allow a geologist to submit such information to the existing geologic description database. The user must indicate his or her name and the geographic area to which the information applies (by assigning a value from the mapextents database). A description category is specified (e.g., geotechnical or hydrologic), and one or more geologic units can also be designated. The anecdote is entered as free text.

Using mapextent methods, anecdotal descriptions can be searched from the geologic map interface along with published descriptions. If the user's map view overlaps the coordinate extent assigned to a particular description, it will be returned in a query. There may be concerns about making information of this kind available to the public, since it has not been formally reviewed. It is important, therefore, to advise users about the nature of the data.

REFERENCE

Weisenfluh, G.A., Curl, D.C., and Crawford, M.M., 2005,The Kentucky Geological Survey's Online Geologic Map and Information System, in D.R. Soller, ed., Digital Mapping Techniques '05—Workshop Proceedings. U.S. Geological Survey Open-File Report 2005-1428, p. 5-10, available at http://pubs.usgs.gov/of/2005/1428/weisenfluh/index.html.

The Challenges and Benefits of Distributing Digital Data: Lessons Learned

By Kenneth R. Papp[1], Susan Seitz[1], Larry Freeman[1], and Carrie Browne[2]

[1]Alaska Division of Geological & Geophysical Surveys
3354 College Rd.
Fairbanks, AK 99709-3707
Telephone: (907) 451-5039
Fax: (907) 451-5050
e-mail: {ken.papp, susan.seitz, larry.freeman}@alaska.gov

[2]Formerly with the Division of Geological & Geophysical Surveys

INTRODUCTION

The Alaska Division of Geological & Geophysical Surveys (DGGS) has been producing geologic maps using a Geographic Information System (GIS) since 1983 (Davidson, 1998). The maps, reports, and informational publications produced by the DGGS are widely utilized by oil, mining, and resource-based companies, as well as consultants, universities, schools, government agencies, scientists, and private individuals. These users have become more technologically savvy over time, and as a result, user requests for digital data in addition to or in lieu of paper reports have grown exponentially. Since 1983, the DGGS has provided several web-based digital geologic data-distribution tools to accommodate the needs of its users (DGGS Staff, 2005), including a database-driven publications query page (http://www.dggs.dnr.state.ak.us/pubs/pubs.jsp), a geochemistry search engine (http://www.dggs.dnr.state.ak.us/webgeochem/index.jsp), and a geologic map indexer (http://maps.akgeology.info/). Currently, the DGGS provides users with digital versions of its maps and reports as Portable Document Format (PDF) files. The raw digital data that generate each map can be burned to a CD or DVD and purchased for a small fee through a general order process. At the time of this writing, a focused effort is underway to upgrade the DGGS web site to provide users with the digital data[1] used to create the state survey's geological and geophysical maps. This paper discusses the challenges and benefits of distributing digital data on-line.

PROJECT BACKGROUND

The primary goal of digital data distribution is to make the data available to the widest possible user audience in formats that are easily adaptable to typical end-user systems. The DGGS Digital Data Distribution (D3) Project is the result and distillation of a series of discussions convened in 2005 in response to numerous public requests for digital geologic map data and the need to fulfill digital data delivery requirements of some projects. The project provides end-users with a means by which to acquire all of the digital datasets used to create, "on the fly," DGGS geological and geophysical maps in the form of ESRI shape files, raster images (i.e. GeoTIFF), various other data, and metadata as compressed files for download via the World Wide Web. The scope of the D3 project includes the following: (1) Enhance the current publications pages to distribute compressed digital data packages, (2) develop a secure, internal application that will allow DGGS staff to create, on-line and off-line, packages for distribution to the public, (3) develop a means by which to insert metadata file elements into the database, (4) modify database structures as needed for application design, and (5) write accurate documentation for project process steps and changes made to the database.

The process work-flow for the distribution process is shown in Figure 1. The first step will mainly involve "cleaning up" the GIS data, ensuring that each dataset has valid metadata, and loading the metadata into the DGGS database (Freeman, 2001a, 2001b). Once this step is completed, all of the digital data files that comprise each publication (project file) will be archived and indexed into the DGGS database system, creating a record of the subsequent distributable dataset. Steps 2 – 7 will be accomplished by providing the authors with a secure, internal, web-based application that will allow users to index their digital geospatial data files and organize them into "packages" according to publication number, dataset, and then data format type. The GIS Manager will then review the dataset packages for data quality purposes (Step 7). Finally, the distribution package will be published to the

[1]Note: Underlined words are defined in Appendix A.

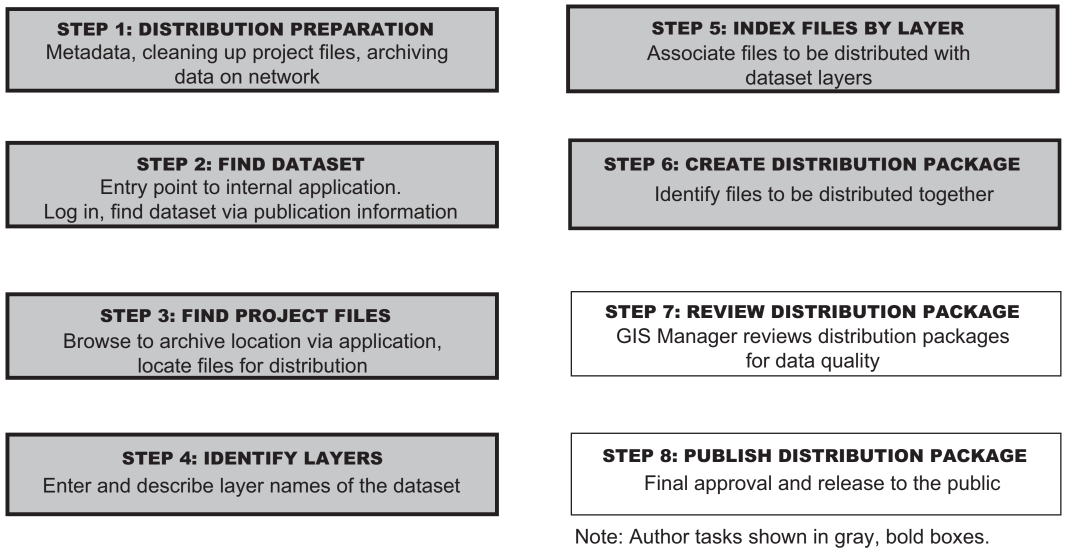


Figure 1. General workflow diagram for the DGGS data distribution process.

DGGS Web site and available to a wide array of end users upon final approval by the DGGS Director (Step 8).

PROBLEMS THAT NEED SOLVING

The old saying “Garbage In, Garbage Out” is certainly relevant in this case, and the process of finding and cleaning up the raw digital data can be daunting and consume a large portion of the D3 project’s resources. Once the data files for publications have been located, recovered, updated, and stored in a central location, a decision needs to be made regarding the data format that will be distributed to the end user. A project of this type and magnitude also requires a well thought-out process work-flow and flexible database structure to store and manipulate the data. Moreover, internal policies and procedures must be created so that all those involved understand the project’s purpose and goals, as well as the assigned roles of staff members and what is expected of them throughout the entire process. Once the data are published on-line and available to the end user, several steps must be taken to ensure that the customer will be satisfied not only with the data, but also the experience he or she will have in *obtaining* the data.

Archiving the Data Pile

The initial step of preparing and archiving the data is similar to taking an inventory of all items in a warehouse. Many organizations have dealt with this issue and have had to make “executive decisions” regarding unknown aspects of legacy data after project managers or veteran geologists retire or leave (Steinmetz et. al, 2002). Moreover, Steinmetz et al., (2002) pointed out that, within the petroleum industry, approximately 60 – 80% of a geoscientist’s time is spent searching for data, while the remainder is spent organizing and analyzing the data. Ensuring that data files are properly cataloged and archived should therefore be a critical priority for any organization that provides data to the public. Documentation and ensuring data quality for legacy datasets are crucial in order to make the datasets meaningful and usable. Over the past several years, the DGGS has participated in the interagency Minerals Data and Information Rescue in Alaska (MDIRA) project. The MDIRA project has allowed the DGGS to overcome many challenges regarding its data archive, specifically database design (Freeman, 2001a, 2001b), restructuring and archiving legacy data by developing an appropriate geologic map model (Freeman and Sturmann, 2004), and writing and organizing metadata (Browne et al., 2003).

At the time of this writing, the total size of geologic geospatial data in the DGGS central server archive is approximately eight times greater than what it was in 2001, which is substantial, given that the average size of a single GIS layer file is on the order of tens to hundreds of kilobytes. There are also many large airborne geophysical datasets, personal databases, and datasets lingering on PCs that have yet to be archived. It is, therefore, the responsibility of project managers and authors to ensure that data files are not lost as tenured staff members leave the organization. In this day and age, the DGGS geologist is expected to solve complex geological problems in the field and in the office, gather information, process data, create a geologic map, and, additionally, archive and document all of the data associated with the project according to current division and FGDC standards. The time required to manage the magnitude of data associated with a given publication often dwarfs the time needed to analyze, understand,

and publish the data. Documentation of many geospatial datasets has been neglected because of geologists' need to initiate new mapping projects. Our hope at the DGGS is that the D3 Project will provide authors and geologists with an effective user interface that would allow them to manage and publish their data more efficiently, granting the geologist more time to "do geology."

What Should We Distribute?

The goal of the D3 Project is to provide datasets in formats that are easily adaptable to typical end-user systems. On the basis of the DGGS staff comments gathered during the project-planning phase, there will be two different file groups for data distribution (Table 1): digital data files and digital data product. After considering the two options, the DGGS decided to use "digital data files" in widely accepted data formats as the minimum standard for all forms of digital data distribution. Providing data to the DGGS's customers in the native data environment format is not the primary goal; however, it is an author-selected option for over-the-counter distribution (see below). Although providing all supporting native dataset environment files with the digital dataset may provide more information with end-user appeal, it is not the standard delivery we recommend. Key reasons for not choosing "digital data products" include: (1) the end-user will require the native software to use the data, (2) greater data liability (e.g. the misuse or misinterpretation of conclusions made by the DGGS), (3) high obsolescence risk, and (4) the need for special knowledge to distribute and use the data (e.g. data/software compatibility and end-user familiarity with the software). Examples of the digital data types distributed by the DGGS are shown in Table 2. These file formats include the most basic data format types that are capable of being adapted and used by a larger end-user audience. Because some data format types are technically proprietary formats (i.e. ESRI shape files, Microsoft Access databases, GeoSoft grid files), providing them in "more generic" formats would be unreasonable due to prohibitive file sizes or the complexity of the common format (i.e. distributing relational databases as ASCII, comma-delimited text files).

The proposed methods of data distribution include provisions for both over-the-counter and on-line distribution. The goal is to make each method distribute the same digital data, but in a different package. Custom distribution orders are always available.

Over-The-Counter Distribution

Digital data will be distributed "on-demand" on transportable media such as a CD-ROM for the over-the-counter (off-line) method. In this case, the publication number is the basis for distribution. Each CD-ROM will sell for $10 to cover the cost of the media, plus applicable postage/shipping costs. Over-the-counter contents could include:

- "ReadMe" file comprised of the table of contents, general information, and instructions in the use of the data or data product (standard)
- Metadata (HTML, ASCII text, and XML formats) (standard)
- PDFs of maps and reports for the publication (standard)
- Digital data files; format depends on data type as per Table 2 (DGGS Standard)
- Native dataset files, where different than digital data and if centrally stored and cleaned up (at author's option)
- Native dataset environment files, if centrally stored and cleaned up (at author's option).

Table 1. Comparison of two categories of data distribution.

Digital Data Files		**Digital Data Product**	
Pros:	Cons:	Pros:	Cons:
1. Simple to distribute 2. Wide audience 3. Easy to index 4. Consistent between projects and publications 5. Minimizes obsolescence 6. Smaller number of files 7. Not dependent on directory structure	1. User processing required before use 2. Annotation may be missing or in metadata 3. Requires export from native dataset 4. Larger file size	1. Data immediately usable/viewable 2. No file conversion (in native environment) 3. Full annotation 4. Contains all built-in logical relations 5. Often used by authors - formats already exist	1. Dependent on directory structure 2. Requires native data environment 3. Larger chance of data liability 4. High obsolescence risk 5. Requires more indexing 6. Difficult to manage 7. May require special knowledge to distribute and use
ASCII (comma or tab delimited), ESRI Shape, Geo-referenced TIFF, MSExcel, ArcExport		Digital data in native data environment (e.g., Geodatabase) *AND* supporting information like symbols, fonts, workspace files, base maps, etc.	

Table 2. Types of digital data formats.

Examples of digital data types	Digital Data Files (DGGS Standard)	Native Data Set Files	Native Data Set Environment Files
Tabular data	ASCII comma, tab delimited	Excel, Lotus 123, or other spreadsheets	NA
Vector data	ESRI shape files	ESRI files, geodatabase, MapInfo tab files	MapInfo workspace, ESRI Map document, fonts, symbol sets, shade sets, etc.
Raster data	TIFF and world file	TIFF and MapInfo tab files	
Grid data	ASCII comma or tab delimited, Geosoft grid or ESRI grid (size of ASCII files may be prohibitive)	ESRI grid files, MapInfo vertical files, ER Mapper grid files	
Relational databases	Native formats accepted here (i.e. MS Access), otherwise ASCII comma, tab delimited	Access, MySQL, or FileMakerPro database	Report, query or data entry documents (HTML, MSWord, Java, PSP, or ASP)

Furthermore, the data storage for the distribution files will use the existing directory structure (Freeman and Sturmann, 2004). All files distributed will be indexed in the database such that they can be located and copied onto distribution media on an "as needed" basis.

On-line Distribution

Digital data will be distributed on-line, free of charge, in compressed files to reduce volume and increase download speed. Compressed files will allow the DGGS to package metadata and other necessary documentation with the selected data as well as preserve any required internal file structure. Each compressed file will contain an individual digital dataset and metadata file, and will be listed with documentary information as an extension to the existing DGGS Publication Web Page for any given publication.

1. Each digital dataset distributed on the Web will display an abstract and have links to:
   - Compressed file containing a digital dataset as digital data files and metadata
   - Metadata file (including code set documentation) for the digital dataset
   - A link to the "ReadMe" File
   - Decompression instructions
2. Information about availability of over-the-counter "data on disk" will be included on the publication page with the following information:
   - Ordering instructions
   - A copy of the "ReadMe" file which includes the disk's table of contents

The Data are Out There, Now What?

With the data files archived, indexed, and bundled into distinct datasets, and metadata written, it may be tempting to think that the job is done. At this point, however, certain aspects of the project are becoming relevant. For example, project managers and geologists must review the final layout of the publication page and datasets before they are officially posted to the Web, despite any previous quality assurance testing.

We are describing a major change in the functionality of the DGGS Web site. These changes will affect users and cooperators, which warrants some sort of notification. It would be beneficial to identify key end-user groups and notify them via the Web site itself, e-mail lists, monthly reports, meetings, or phone calls. Once end-users are aware of the new data-distribution service, it is imperative to provide effortless feedback methods with which these users can comment on data quality and ease of use, and submit suggestions. Similar to the open-source software community, the multitude of end-users are relied upon to find any remaining "bugs" in the system. Moreover, the DGGS will utilize database log files and web statistics to identify the most "popular" datasets and get a better understanding for what information is in demand.

LESSONS LEARNED

It was imperative when designing the D3 Project that the data distribution methods for DGGS staff were consistent and clearly stated. The D3 Project designers met with geologists and project leaders to discuss the distribution work-flow, user interface, responsibility assignments, and details of particular types of datasets and archival strategies. The data distribution process should also be flexible to meet changing expectations and technical requirements of end-users. For example, breaking up the publications into several on- and off-line datasets provides flexibility and benefits those with small bandwidth or no Internet access.

Prior to distributing data to the public, an in-house inventory of existing data serves to identify which data are at risk. This process benefits both the distributor and the end-user by ensuring that the data adheres to current documentation standards, and by securing the data on more reliable media. Many agencies take the risk of storing and distributing data in proprietary data formats that may soon become obsolete or unreadable. With regard to such a risk, one has to ask, "Which, if any, software will be available 5, 10, or 20 years from now that can read the data?", and "When might the data become legacy data?"

In theory, data are always becoming legacy in status when software vendors upgrade their program packages, hardware becomes obsolete, and geologic maps are updated. Many agencies invest a large amount of time programming and creating scripts in the current software version, only to find that those scripts are worthless in the next program release. Similarly, storing precious data on only one type of archival media can be a terrible mistake. It is, therefore, up to project managers and authors to know when valuable data may be at risk and establish a legacy data recovery plan to prevent future data loss. Implementing a project such as this forces the agency to "clean house" and index valuable data.

Everyone involved with these kinds of projects must understand that documentation and data-quality information for every dataset are required. As a result, end-users will get consistent, quality data that are well-documented, which will allow them to have access to the information they need to use the dataset. If a user of a given dataset cannot find its documentation, he or she will more than likely (1) not use it, (2) attempt to use the dataset without proper guidance and understanding, or (3) use the dataset incorrectly or inappropriately. If project managers and authors take the time to document their data soon after it is created, the painstaking process of going back through tens or hundreds of datasets (some 20 years old), contacting retired staff members, and guessing about the details of a publication can be avoided. Moreover, by automating distribution methods to the greatest extent possible, the data can be delivered on demand. Since the freely provided data are already in digital form, easily searchable, well documented, and organized by dataset, users can focus on merging the data into their own projects and spend more time on analysis and understanding the implications of their scientific data and observations.

REFERENCES

Browne, C.L., Freeman, L.K., and Graham, G.R.C., 2003, The Alaska Division of Geological & Geophysical Survey's Metadata Policy Development and Implementation, in Soller, D.R., ed., Digital Mapping Techniques 2003—Workshop Proceedings: U.S. Geological Survey, Open-File Report 03–471, p. 201–208, available at http://pubs.usgs.gov/of/2003/of03-471/browne/index.html.

Davidson, Gail, 1998, Can we get there from here? Experiences of the Alaska Division of Geological and Geophysical Surveys, in Soller, D.R., ed., Digital Mapping Techniques 1998—Workshop Proceedings: U.S. Geological Survey, Open-File Report 98-487, p. 13–15, available at http://pubs.usgs.gov/of/of98-487/davidson.html.

DGGS Staff, 2005, Alaska GeoSurvey News: Alaska Division of Geological & Geophysical Surveys Newsletter 2005-2, 8 p., available at http://www.dggs.dnr.state.ak.us/pubs/pubs?reqtype=citation&ID=14595.

Freeman, L.K., 2001a, The DGGS Geologic Database: Putting Geologic Data Modeling into Practice: Alaska Division of Geological & Geophysical Surveys, Alaska GeoSurvey News, v. 5, no. 3, 3 p., available at http://www.dggs.dnr.state.ak.us/pubs/pubs?reqtype=citation&ID=14588.

Freeman, L.K., 2001b, A Case Study in Database Design: The Alaska Geologic Database, in Soller, D.R., ed., Digital Mapping Techniques 2001—Workshop Proceedings: U.S. Geological Survey Open-File Report 01-223, p. 31–34, available at http://pubs.usgs.gov/of/2001/of01-223/freeman.html.

Freeman, L.K., and Sturmann, A., 2004, Progress Towards an Agency-Wide Geologic Map Database at Alaska Division of Geological & Geophysical Surveys, in Soller, D.R., ed., Digital Mapping Techniques 2004—Workshop Proceedings: U.S. Geological Survey Open-File Report 04-1451, p. 9–14, available at http://pubs.usgs.gov/of/2004/1451/freeman/index.html.

Steinmetz, J.C., Hill, R.T., and Sowder, K.H., 2002, Digital Archives and Metadata as Mechanisms to Preserve Institutional Memory, in Soller, D.R., ed., Digital Mapping Techniques 2002—Workshop Proceedings: U.S. Geological Survey Open-File Report 02-370, p. 171–176, available at http://pubs.usgs.gov/of/2002/of02-370/steinmetz.html.

APPENDIX A

(Description of Terms)

Custom distribution: A custom distribution is a combination of data or data derivative that has not already been generated via the publication process. This may include requests for data reprojections, file format conversions, combining GIS layers from multiple projects or publications, statistical or spatial analyses, and excessively large amounts of data.

Dataset: A unique group of data that acts as a component of the publication. Examples include vector geologic features (i.e. bedrock, surficial, hazard polygons/lines/points), geochronology (i.e. spreadsheets, ASCII .csv), DEM data, electromagnetic anomalies, and grid data.

Digital data: Information that is ready for numeric or geographic manipulation with a minimum of conversion or preparation by the customer (e.g. Excel spreadsheets, formatted ASCII files, relational databases, geo-referenced raster files, geo-referenced vector graphics files).

Digital dataset: A logical, thematic, and geographic grouping of data, including any code sets (required to interpret the data). There may be one or more datasets per publication; a metadata document describes a digital dataset. Examples include GIS bedrock geology and spreadsheets that contain geochemical data related to a single publication.

Digital data file: Digital data in a file format that can be used across a wide variety of computing systems and meets the needs of most data consumers (See Table 1). These should be the standard formats that DGGS uses to distribute digital data.

Digital data product: Provides data and supporting information required to view the data in the native dataset environment (See Table 1). An example includes an ESRI Geodatabase and all supporting information like symbols, fonts, workspace files, base maps, etc.

Layer name: The name of the GIS layer, coverage, TAB file, or table name as defined in the metadata by the DGGS metadata extension, *Entity_and_Attribute_Layer_Name* (See Steps 4 and 5, Figure 1). If no layers exist in the metadata, the author may have to create layer names for their dataset within the application for the purpose of indexing their files.

Metadata: Metadata consist of information that characterizes data. Metadata are used to provide documentation for data products. In essence, metadata answer who, what, when, where, why, and how about every facet of the data that are being documented. Metadata written by the DGGS must conform to FGDC standards (http://www.fgdc.gov/metadata/geospatial-metadata-standards). Metadata will be distributed in three file formats to allow maximum readability and usability: Frequently Asked Question (FAQ) HTML, ASCII plain text, and Extensible Markup Language (XML).

Native dataset: Digital data in file formats that were produced by the software that was used to generate and process the digital data; the dataset does not include supporting native environment files (See Table 2). The user of these datasets may need access to the same software version that was used to produce the data.

Native dataset environment: The software operating system, hardware, and supporting files used by the producer to create, view, and process the dataset (See Table 2); it may be specific enough that it could be very difficult to replicate.

On-line distribution: Provides the e-mail and web browser customers with digital data in the form of compressed downloadable data.

Over-the-counter distribution: Provides the phone, mail, and walk-in traditional customers with digital data on some media (e.g., CD-ROM).

Project file: Any file found within the publication or project directory located in the DGGS directory structure on the central fileserver.

The Alabama Metadata Portal: http://portal.gsa.state.al.us

By Philip T. Patterson

Geological Survey of Alabama
420 Hackberry Lane
P.O. Box 869999
Tuscaloosa, AL 35468-6999
Telephone: (205) 247-3611
Fax: (205) 349-2861
e-mail: PPatterson@gsa.state.al.us

INTRODUCTION

In recent years federal, state, and local government entities in Alabama have made substantial investments in the collection, management, and use of geospatial data. However, there has been no large scale effort to share data effectively and efficiently. The result was unnecessary expenditures in redundant data creation.

Most Alabama Geographic Information Systems (GIS) users currently have broadband internet access. The increased network connectivity and high data-transmission rates have produced the expectation that large amounts of data can be accessed instantly. This demand for data access has motivated the Alabama Emergency Management Agency (AEMA) and the Geological Survey of Alabama (GSA) to collaborate in developing the geospatial data portal, which allows cooperators and users to search for, discover, and access geospatial data (GSA, 2006).

BACKGROUND

Before starting the project, extensive research on a variety of data delivery options was performed. The majority of the options were related to data clearinghouses, which are mainly useful for specific types of static data like imagery, civic boundaries, center lines, etc. However, the data delivery website to be built would not be intended for static data alone. The need was to build a robust compilation of all different types of vector and raster data, ranging from general datasets to obscure data specific to individual projects. Also long-term administration responsibilities for this type of complex compilation site were a concern for GSA. Eventually, the grant for site development would end, and GSA would have to support managing and updating the site from internal resources.

With support from Environmental Systems Research Institute Inc. (ESRI), we addressed this concern with a modified out-of-the-box application using open-source web applications in conjunction with ArcIMS, ArcSDE, and an underlying database management system (DMS). The resulting site provides the functions of a clearinghouse for general data and a search engine for unique data. It also offers semi-automated administration, which allows users, as well the administrator, to manage the site. This solution is ideal in addressing the data delivery goals and the long-term administration concerns posed by this project.

CONNECTION

This search engine and download site provide the framework for a mutual geospatial user community of organizations and stakeholders that facilitates discovery, sharing, and delivery of GIS content and services. The portal also facilitates the organization of content and services such as directories, search tools, community information, and support resources applications.

The underlying structure of the portal is a three-part generalized connection as follows (Figure 1): (1) the portal connects to a data provider's metadata library, which grants users the rights to publish specified metadata records to the portal's online catalog; (2) the data user connects to the portal's search option to locate data using the portal's search engine without physically browsing through the stakeholder's data; and (3) the data users will connect to the data provider for download, data captures, or the identification of the data resource.

By storing only metadata records in our catalog, we have the ability to index a large amount of virtual data, and more importantly, the GSA and AEMA will not have to store the physical data. Our goal is to automate the tasks of data discovery and distribution so that once portal connections are complete, minimal maintenance is required from the hosting agency.

ARCHITECTURE

A portal is essentially a master web site, which is connected to a web server and contains a database of metadata information about geographic data and services. The services are exposed as web applications using open

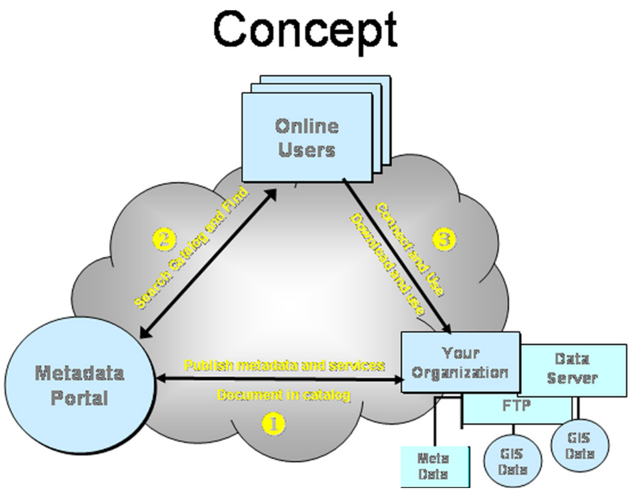


Figure 1. Generalized data partnership and user connection concept (modified from ESRI, 2004).

source environments (Tomcat, Java, html, http, xslt, xml, and jsp) to provide a user-friendly and visually appealing interface.

The architecture of the metadata server, which connects to all indexed metadata records, relies on three existing ESRI products: ArcIMS, ArcGIS, and ArcSDE. The ArcIMS provides the framework and architecture on which the metadata server runs. The ArcGIS ArcCatalog application serves as an authoring and publishing tool. The ArcSDE stores published metadata in records inside a relational database (ESRI, 2004). ArcIMS introduces a new approach to serving map products over the internet through a Java-based application management environment that includes mapping services and map design tools to support a variety of internet map services (ESRI, 2004). Main components associated with the ArcIMS communication architecture and web applications are identified in Figure 2.

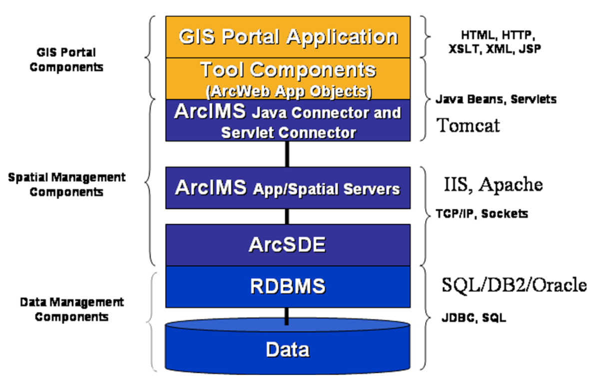


Figure 2. GIS software environment (Modified from ESRI, 2004).

THE PORTAL'S ONLINE INTERFACE COMPONENTS

Home Page

The home page shown in Figure 3 is the access point for all online components, and it provides quick access to the most popular data applications. From the home page, a user can do a basic keyword search, navigate to the map viewer, find help information, and access the quick links to downloadable data, GIS projects and services, and GIS resources.

The home page is also where users login to their accounts. A user account is not necessary to access the portal, but it increases user capability and enhances functionality. There are five distinct user levels of the portal based on a top-down hierarchy; that is, higher level users can do everything a lower level user can do. From lowest to highest, these include:

1. Anonymous users can be anyone. These users have the ability to browse the site and use three basic

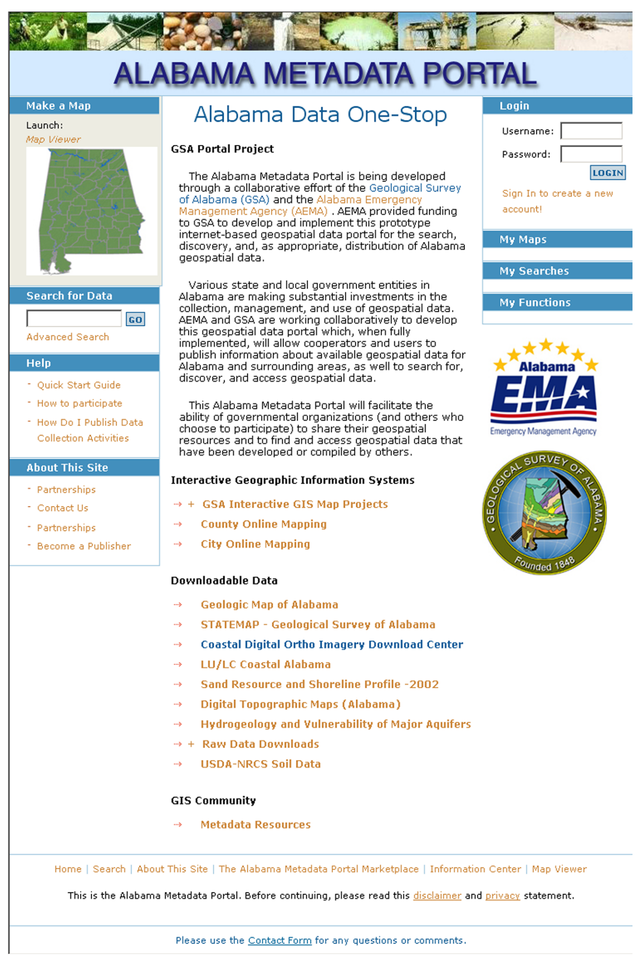


Figure 3. Example of the portal homepage.

online components: home page, map viewer, and search page.
2. Public Users have the ability to save their created maps from the map viewer and save their data searches, which will be available on the users' home page.
3. Publisher Users have the ability to create, publish, and manage their metadata online.
4. Channel Managers have the ability to create and publish a quick link on the home page.
5. Administrators check metadata for accuracy, batch-upload metadata, harvest publisher metadata, and manage users.

Map Viewer

The portal map viewer shown in Figure 4 is a mapping application that allows users to view one or multiple internet map services at the same time in their web browser. Access to selected federal, state, and local Web Map Services (WMS) using the "add service" menu is provided, but this limited number of services can be expanded by entering other map server URL addresses to access other WMS available online. Viewing internet map services through the portal map viewer allows users to:

- add map services from the portal and other map servers
- display one or multiple map services in a single map view
- set the transparency of map services for overlaying multiple images
- turn map layers on or off within a map service
- find latitude/longitude anywhere in the state for automatic navigation of the map
- find street addresses in the state for automatic navigation of the map
- identify attribute information about features in a map service.

The portal is not limited to just ArcIMS WMS; it also supports several specifications and services of the Open Geospatial Consortium (OGC). The OGC is a non-profit, international, voluntary consensus standards organization that is leading the development of standards for geospatial and location-based services (OGC, 2006). The portal supports the following specifications from the OGC:

- Web Mapping Services versions 1.0, 1.1, and 1.1.1
- Web Feature Services version 1.0.0
- Web Coverage Services version 1.0.0

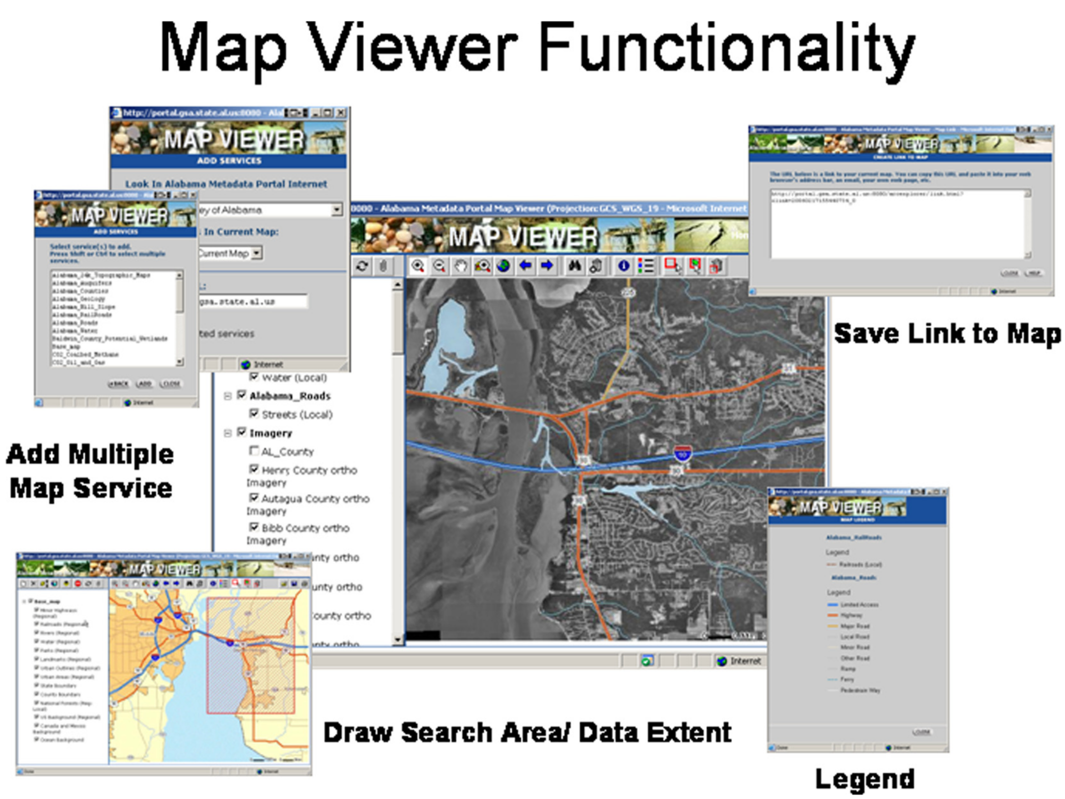


Figure 4. The portal's map viewer.

- Web Map Context Documents version 1.0.0
- Geographic Markup Language versions 2.0 and 3.0 (when approved)
- Open GIS Location Services version 1.0.

Search Function

The search page shown in Figure 5 is the tool for searching and discovering the metadata of content offered by many publishers of the Alabama Metadata Portal. The search page allows users to specify the geographic extent, keywords, content type, or content theme criteria to find matching metadata of map services, data, maps, web services, activities, or documents published in the Alabama Metadata Portal. Users can search the portal by defining "where" they would like to search, "what" in the state they would like to search, and "when" they would like the content they are searching for to have been created or updated. Users only need one parameter for a simple search; however, each additional parameter helps to narrow or retrieve a search.

THE METADATA PUBLISHING FUNCTION

The importance of writing good metadata is difficult to communicate to potential publishers of the portal. The success of the connection in Figure 1, however, is based on accurate and current metadata. Metadata describes the who, what, when, where, why, and how questions about the data, which gives users the knowledge to decide whether the data is appropriate for their desired application. Writing good metadata also mitigates the overall burdens and cost of data maintenance. The standards for including metadata records in the portal are the Federal Geographic Data Committee's (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM).

There are three user levels that have metadata administration: Publisher, Channel creator, and Administrator. The administration of metadata includes the ability to create, manage, and add metadata to the portal. There are three options to make metadata records available for search in the portal. The first option is to publish a

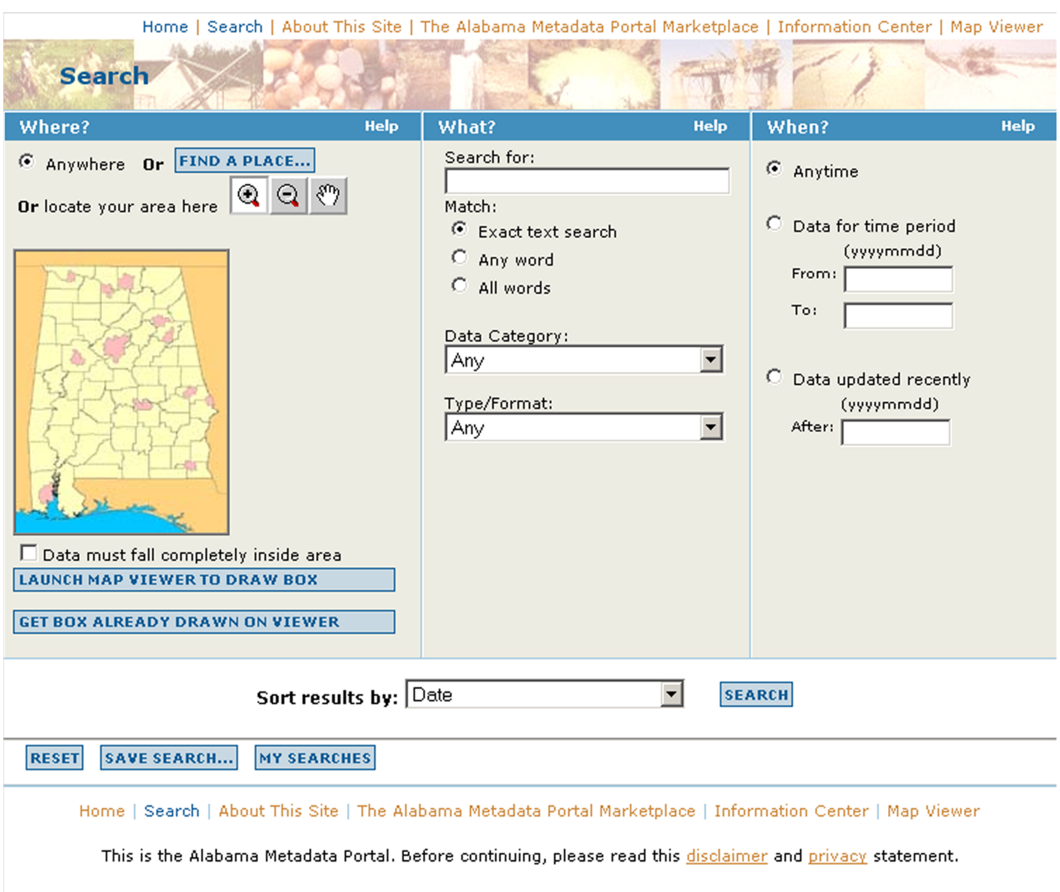


Figure 5. The portal's advanced search page.

metadata collection to a metadata repository where the portal can harvest it. The second option is to upload an individual or batch Extensible Markup Language (XML) formatted metadata record to the portal. The third option is to create a metadata record online using the portal's metadata creation tool.

Metadata Harvesting

Metadata harvesting is a self-regulated, scheduled process for collecting new and updated metadata from various metadata collection libraries. The process of harvesting allows the portal to synchronize its metadata repository with the publisher's metadata catalog. If publishers participate in metadata harvesting, any updates made to their local metadata collection will be updated in the portal during the next harvesting session. Currently, the portal can harvest FGDC-compliant metadata from three different types of harvesting protocols: Z39.50 metadata clearinghouse node, ArcIMS metadata service, and Web Accessible Folder.

Metadata harvesting in the Alabama Metadata Portal is performed in three steps as shown in Figure 6:

1. Harvesting: Based on harvesting protocol specified at the time of registration, the portal will connect to the user's local metadata repository and retrieve all new and updated metadata records.
2. Validation: During validation, the portal administrator examines each metadata record to confirm that minimum portal requirements are met. Records that are rejected are sent back via e-mail with a list of invalid fields that need to be added. The records will not be added until the metadata record is corrected and revalidated.
3. Publishing: All successfully validated and accepted metadata is published in the portal database. Once the metadata is published, it is searchable through the portal's search interface by all users.

Direct Metadata Upload

If users do not have access to any of the metadata distribution server protocols as described above, they can upload their XML-formatted metadata records directly to the portal. A metadata publisher can, through the online administration tool, add and manage metadata on their homepage. Selecting the "Upload Metadata" button, users can upload individual metadata records saved on their local computer. These records will be validated and either rejected or published in the same process as metadata harvesting. A drawback to the direct upload option is that uploaded published metadata is not linked to the local metadata repository. That is, updates to a local metadata record must be uploaded or manually changed because they are not automatically updated by the portal when the user updates local records.

ArcCatalog Direct Metadata Upload

Batch uploading of metadata records directly to the portal's metadata IMS service is possible if the user is using ESRI's ArcGIS suite. Through ArcCatalog, the user will directly connect to the portal's ArcIMS metadata server; a metadata publisher account name and password must be specified. With this connection to the portal in place, the users can drag and drop their folder of meta-

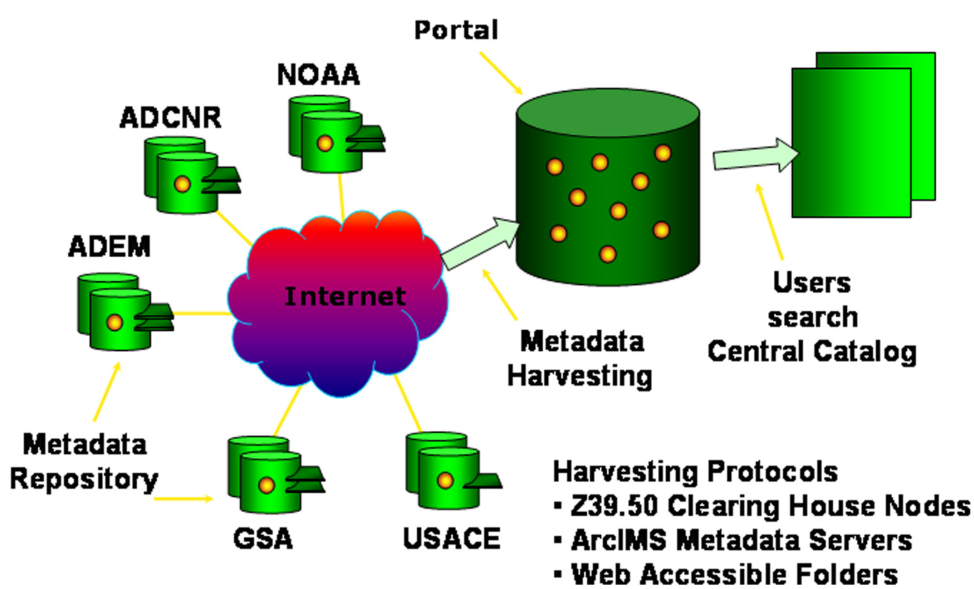


Figure 6. Diagram of the harvesting process(Modified from ESRI, 2004).

data records into the Publish Metadata Service. An added benefit to this drag-and-drop method is that the metadata record is validated automatically and displays an error message for all incorrect field values. The drawback of this method is the same as the direct upload option where uploaded published metadata is not linked to the local metadata repository as they would be with harvesting. Updates to a local metadata record must be uploaded or manually changed because they are not automatically updated by the portal when the user updates local records.

Metadata Direct Entry

The user might not have access to metadata creation or editing software, or may have very few records to contribute to the portal. If this is the case, the user can utilize the metadata creation tool provided on the home page. Users will login to their account and find the "publish online form" button under the "My Function" section. This button will take users to an online form designed to assist users in the development and production of FGDC metadata quickly and efficiently. The form provides the users with drop menus, fields that are required (indicated by *), as well as help definitions and suggestions for each of the requested metadata fields.

The minimal compliance of the direct entry method provides only the elements necessary for data discovery and is only moderately functional to users searching for data. The direct entry is a means by which to encourage users to write metadata in the hope that they will see its importance and progress toward creating a comprehensive FGDC-compliant record in the future. By using the online creation tool, the metadata will be stored only in the portal, and all updates must be made through the portal.

CONCLUSION

Data download sites and web applications have dramatically improved the GIS productivity. To complete jobs faster, it is critical that the GIS community share data effectively and efficiently: the portal is a powerful tool that benefits all users and addresses these needs. Faster discovery of specific datasets and projects, data access to download sites and use in the online Map Viewer, lowering of data costs by reducing the redundancy of data, comparison of multiple providers to find data that suits their needs, and improvement of data quality and coverage with a constant updating of agency metadata are a few benefits available through the portal. More importantly, the portal heightens the visibility of participating organizations by displaying the quality and quantity of their data offerings, which is an indication of their GIS capabilities. This allows a better understanding of how an organization could partner for future projects or initiatives.

The first 18 months since the activation of the Alabama Metadata Portal, there were 378,225 total domain hits, which represent 16,197 visits by 5,544 unique users (unique IP addresses) shown in Figure 7. We estimate that each return user has viewed an average of 68 pages. This current assessment shows the effectiveness of the Alabama Metadata Portal and the public's interest in accessing the data provided. It is important to note that the portal initiative is by no means the sole solution in producing an integrated GIS community; the portal represents a fundamental step moving Alabama into the next generation of GIS productivity.

REFERENCES

Environmental Systems Research Institute, Inc. (ESRI), 2004, GIS portal technology, an Environmental Systems Research Institute, Inc. White Paper: Redlands, California, 9 p.

Geological Survey of Alabama (GSA), 2006, Geological Survey of Alabama Portal Project: Alabama Metadata Portal, accessed 01 Sept. 2006 at <http://portal.gsa.state.al.us>.

Open Geospatial Consortium, Inc. (OGC), 2006, Welcome to the OGC website: accessed 01 Sept. 2006 at <http://www.opengeospatial.org/>.

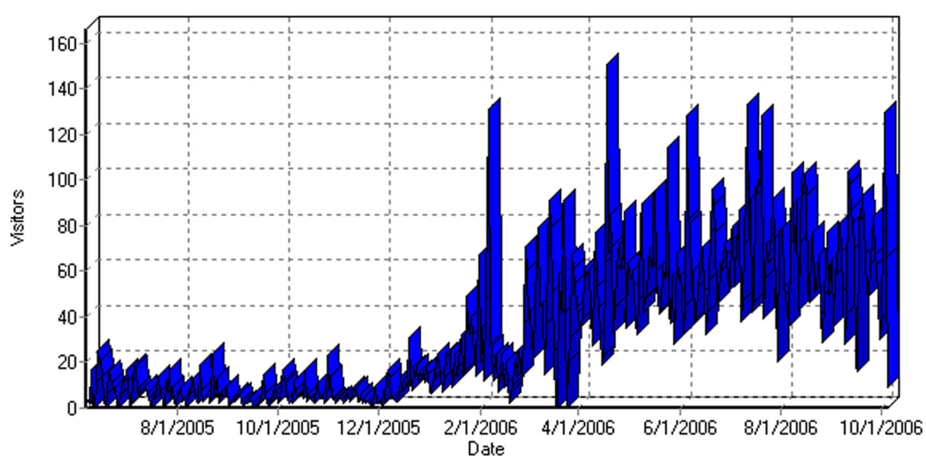


Figure 7. Daily hits on the portal from 07/01/2005 – 10/01/2006.

3D Geological Modeling: Solving as a Classification Problem with the Support Vector Machine

By Alex Smirnoff, Eric Boisvert, and Serge J. Paradis

Geological Survey of Canada
GSC-Quebec
490 de la Couronne
Quebec, Canada G1K 9A9
Telephone: (418) 654-3716
Fax: (418) 654-2615
e-mail: alsmirno@nrcan.gc.ca

ABSTRACT

The process of creating multi-unit 3D geological models by successive unit interpolation may be tedious and time-consuming. Here, we propose to automate this procedure through presenting the problem as a classification task and solving it simultaneously with the Support Vector Machine (SVM), a method known from the field of artificial intelligence. Experiments with various input data and kernel parameters demonstrated that the SVM has great potential in 3D reconstructions from sparse geological information. An extended version of this paper has been accepted for publication in "Computers and Geosciences" (Smirnoff et al., 2008).

INTRODUCTION

Often, geologists are faced with a variety of diverse information that requires generalization and analysis. 3D modeling software packages such as Gocad® of Earth Decision Sciences have proven an excellent means for data presentation and interpretation. The procedure normally requires reconstruction of individual geological units using surfaces interpolated from control points with subsequent fusion of these units into a single model. The popular interpolation techniques include Inverse Distance Weighting (IDW), Discrete Smooth Interpolation (DSI), and various flavors of kriging preceded by semi-variogram analysis.

The above procedure can easily become a tedious and time-consuming task when a complex geomodel is considered. In addition, the traditional interpolation techniques assume reasonable areal coverage of the input data. Therefore, there is a strong need for an algorithm that would automate the process of model creation even in cases when only a few pieces of information on regional geology, (e.g., a few sparse cross-sections) are available. Finding such an algorithm and testing its performance on available data sets was the objective of this study.

Here, we propose the use of the Support Vector Machine (SVM), a tool routinely applied in the field of image analysis and pattern recognition. The SVM is becoming increasingly popular and has been successfully used to solve classification and regression problems in biology (e.g., Noble et al., 2005), hydrology (e.g., Yu et al., 2004), medicine (e.g., El-Naqa et al., 2002), and environmental science (e.g., Gilardi et al., 1999). In this study, we demonstrate that the application of SVM in geology allows sparse data to be efficiently combined in order to reconstruct shape, area, and volume of multiple geological units.

METHODOLOGY

The SVM Algorithm

The SVM algorithm is based on the Statistical Learning Theory developed by V. Vapnik (Vapnik, 1995). It uses a set of examples with known class information to build a hyperplane that separates samples of different classes. In machine learning theory, this is known as supervised learning as opposed to unsupervised learning when no a priori class information is available. This initial dataset is known as a training set, and every sample within it is characterized by features upon which classification is based. Figure 1A demonstrates this for the one-dimensional (single-feature) case. The samples closest to the hyperplane are termed support vectors (filled marks in Figure 1).

In more complicated, non-linear cases, the task of discovering the separator is turned into a linear task by transferring input data into a higher-dimensional space known as the feature space. Figure 1B shows a non-separable one-dimensional data set in the input space. The problem is easily solved through re-mapping data to a higher, two-dimensional space where a linear solution is found (Figure 1C). Functions satisfying certain conditions and known as kernels are normally employed for this

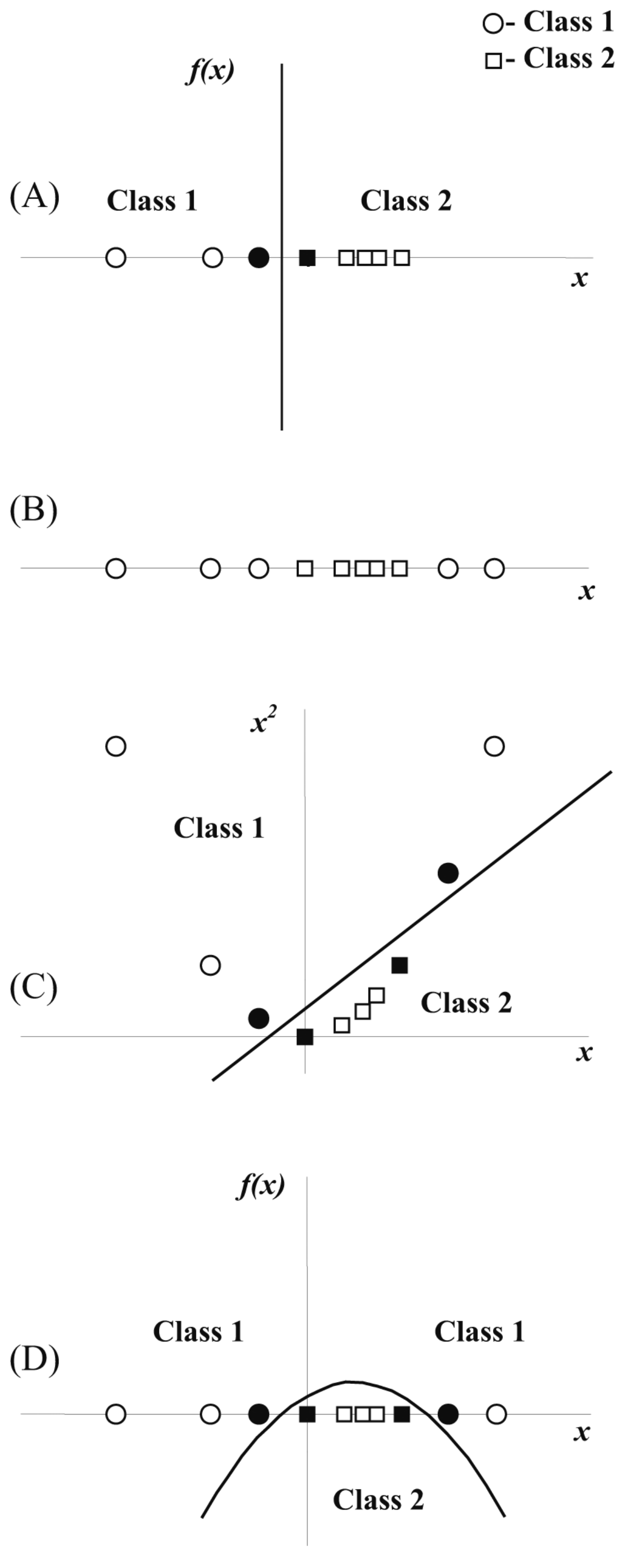


Figure 1. Building the separating hyperplane, in separable and non-separable one-dimensional case. Filled marks represent support vectors. (A) Linearly separable case and decision function in input space; (B) non-separable case in input space. (C) training data re-mapped into two-dimensional feature space using $\phi(x) = (x, x^2)$ and linear solution in this space; (D) solution re-plotted back in input space.

transfer (e.g., Abe, 2005). The solution becomes non-linear when shown in the original data space (Figure 1D).

Once the equation for the optimal classifier is found, new data with unknown class information (test samples) can be classified based on the value of this decision function. Unlike most interpolation methods based on the principle that values at points closer in space are more similar, the SVM is a boundary classification method where the boundary is built based on the initial training set among which only a small number of samples (support vectors) are involved in the final decision making.

The classical SVM task is a binary (two-class) classification. However, a number of methods have been developed to support multi-class classification through various combinations of binary methods such as "one-against-all", "one-against-one", etc. (see Hsu and Lin, 2002 for references). Therefore, the SVM approach is also applicable for models with more than two classes. More detailed descriptions of the SVM algorithm are available from a number of sources (e.g., Cristianini and Shawe-Taylor, 2000; Abe, 2005).

SVM Application to 3D Modeling

To apply the SVM algorithm to our geological reconstructions, we defined the 3D space-partitioning task as a pure spatial classification problem. Three coordinates uniquely describe every point in the 3D reconstruction space. However, only a limited number of those points possess descriptions or class information that can be identified through well drilling, surficial geology mapping, and seismic profiling. The class information describes the geological unit to which each particular point belongs. Therefore, the points with known class labels become samples in the SVM training set, and point coordinates in the three-dimensional space are used as sample features. Once a classification model based on this training set is built, the rest of the points in the reconstruction space can be classified based on their coordinates (features).

We employed one of the many SVM implementations freely available over the internet, namely LIBSVM developed at the National Taiwan University (see Chang and Lin, 2001 for detailed description). As recommended in Hsu et al. (2004), we used LIBSVM with the radial basis function (RBF) kernel, the most general form of kernel resulting in a prediction model controlled by only two hyperparameters, C and γ. A single solution is obtained for every pair of parameters, and it is sensitive to the choice of their values. However, selecting the appropriate values is a dark art normally done on a try-and-see basis.

For multi-class classification, LIBSVM uses "one-against-one" approach, which was shown to be advantageous to other methods for practical use (Hsu and Lin, 2002). In addition, a set of in-house Java utilities has been developed for scaling, validation, and format conversion purposes.

Data

A 3D geomodel created at the Geological Survey of Canada, Quebec, in the course of the Esker/Abitibi project (Bolduc et al., 2005) was used as the reference dataset in all of the experiments. This model is based on surface geology, well, and cross-section data. Six geological units were sequentially interpolated from the above control points using the Discrete Smooth Interpolation (DSI) technique (Mallet, 1989) in the Gocad® GIS.

Experimental Work

The experimental work was designed to perform the following tasks: (1) investigate whether the SVM can be efficiently used in binary geological reconstructions, (2) test the SVM for multi-unit modeling, and (3) examine how the resulting model depends on the RBF kernel parameters used in the reconstruction.

General Approach

The general approach taken in all of the experiments was as follows:

Prediction

- In Gocad®, create a reconstruction space as a set of volume elements (voxet) of the shape representative of study area geometry. The reconstruction space was defined by a voxet with the following number of volume elements (voxels) in each direction: X-110, Y-240, Z-24.
- Add available data to the reconstruction space. The unit type property for each of the six geological units (SVM classes) was transferred from the stratigraphic grid (SGrid) structure representing the reference model.
- In Gocad®, using a DEM, define all voxet nodes above the surface as air or no-data points.
- Define a training set for the experiment. For the remaining ground points, set the unit type property to zero; these are the points that will be later classified by trained SVM.
- Scale coordinate values for the training set between 0 and 1 as recommended in Hsu et al. (2004).
- Using LIBSVM, build a prediction model based on the training set. A single reference set of kernel parameters ($C = 10^4$ and $\gamma=10^2$) previously determined from 2D experiments was used in all reconstructions except the parameter sensitivity tests.
- Scale coordinate values for the points to be classified and classify them using the prediction model created in the previous step.
- Import the point set with predicted class information back into Gocad®, for visualization and analysis.

Validation

- Based on the available reference data, define the validation set and extract it as a set of points with attached class property.
- Test predicted class labels against the validation set to determine how many original points in each class and overall were adequately classified, a measure also known as the recall rate.

Binary Reconstruction

The training set was composed of the Esker/Abitibi model points located on 11 arbitrarily chosen parallel sections oriented along axis X. Points were grouped into two classes as shown in Table 1. The input data statistics are given in Column 4 of Table 1. As seen from the table, the training set was dominated by points representative of Class 2, which combined all model units except the Esker Unit. The validation set was composed of all the remaining model points (not included in the training set). The number of points used for validation in each of the two classes is shown in Column 1a of Table 3.

Table 1. Geological units, SVM classes, and training set statistics for Esker/Abitibi Binary Model. Total number of points to be classified is 371783.

1. Geological Unit	2. SVM Class	3. All Points (#/%)[b]	4. Training Points (#/%)[c]
Esker	1	20300/5.22	995/0.26
Non-Esker[a]	2	368935/94.78	16457/4.23
All Units	-	389235/100	17452/4.48

[a]Non-Esker unit included Rock, Till, Clay, Littoral and Organic units
[b]Number of all class points and their percentage of all model points
[c]Number of training points per class and their percentage of all model points

Multi-Class Reconstruction

The same training points, arranged into six classes corresponding to the six geological units found in the original model, were used to test the SVM capabilities in multi-class classification (Table 2). For training data statistics, see Column 4 of Table 2. This time, bedrock (Class 6) entirely dominated the training set, with organics (Class 1) being the least representative. The validation set also contained information about all six geological units as shown in Column 2a of Table 3.

Hyperparameters Sensitivity Tests and Multiple Parameter-Set Reconstructions

To analyze the sensitivity of prediction results to the values of hyperparameters, C and γ, we used a simple grid search procedure as proposed in Hsu et al. (2004). The grid search was run for the above training set configurations, and the range of parameters scanned by every search was from 2^{-8} to 2^{15} for C and from 2^{-15} to 2^{12} for γ incrementing parameter values by a power of 2. As in previous experiments, the success rate was determined through direct comparison with the validation set extracted from the reference model. We also examined the dependency of success rate on parameter values for the class with the least number of training points (Class 1–Organics). Finally, binary models were built with combinations of parameters drawn from the margins of the reasonable working range. These included low C (2^{-2}) – high γ (2^8), low C (2^{-2}) – low γ (2^5), average C (2^7) – average γ (2^6), high C (2^{14}) – high γ (2^8) and high C (2^{14}) – low γ (2^5).

RESULTS AND DISCUSSION

Binary Reconstruction

The original esker body, training sections, and the results of binary reconstruction with the reference parameter set are shown in Figure 2. Column 1b of Table 3 describes

Table 2. Geological units, SVM classes, and training set statistics for Esker/Abitibi Multi-Class Model. Total number of points to be classified is 371783.

1. Geological Unit	2. SVM Class	3. All Points (#/%)[a]	4. Training Points (#/%)[b]
Organics	1	1210/0.31	48/0.01
Littoral	2	3819/0.97	193/0.05
Clay	3	13295/3.42	628/0.16
Esker	4	20300/5.22	995/0.26
Till	5	15865/4.08	747/0.19
Bedrock	6	334746/86.00	14841/3.81
All Units	-	389235/100	17452/4.48

[a]Number of class points and their percentage of all model points
[b]Number of training points per class and their percentage of all model points

Table 3. Validation data and results for Esker/Abitibi binary and multi-class model. Number of validation points in original model and percentage of points properly classified by SVM.

SVM Class	1. Binary		2. Multi-Class	
	a. Validation Points (#/%)[a]	b. Success (%)	a. Validation Points (#/%)[a]	b. Success (%)
1	19305/5.19	71.50	1162/0.31	18.76
2	352478/94.81	98.87	3626/0.98	37.20
3	-	-	12667/3.41	57.10
4	-	-	19305/5.19	67.65
5	-	-	15118/4.07	45.72
6	-	-	319905/86.05	95.45
All	371783/100	97.34	371783/100	89.87

[a]Number of validation points per class and their percentage of all model points

the validation results. As seen from Table 3, the success rate of SVM prediction is exceedingly high. Especially remarkable results, 98.87%, are achieved in Class 2. In part, this can be attributed to the fact that points of this class entirely dominate the training set. When the SVM cannot classify a point in binary classification, it tends to attribute it to the predominant class. Considering that bedrock points constitute 94.81% of all points that need to be classified (352478 of 371783 as shown in Column 1 of Table 3), it is no surprise that the overall success of prediction achieves 97.34%.

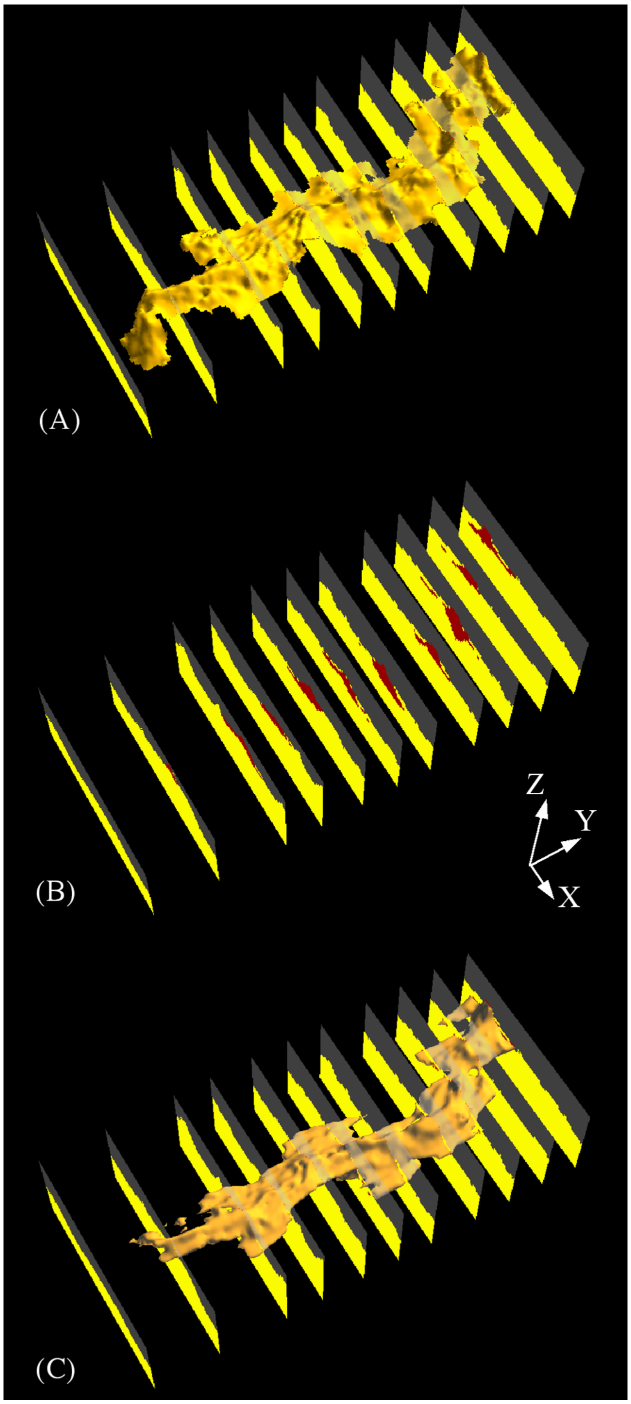


Figure 2. Binary esker reconstruction. (A) Original Esker Unit; (B) training set; (C) reconstructed Esker Unit.

With the above explanation in mind, the classification success in Class 1, which represents only about 6% of the training set, is still as high as 71.50%. This, in our opinion, proves that the SVM can be effectively used for binary (single-unit) reconstructions even with training sets substantially skewed toward one of the classes.

We further analyzed success rate in Class 1 on all model sections where the Esker Unit was present (234 sections). The results are presented in Figure 3. The figure clearly demonstrates that the success of prediction decreases as the distance from a training section increases. As training section # 1 did not intersect the esker body, the success rate on the first 18 sections drops to 0%. Therefore, as could be expected, the overall reliability of prediction is directly proportional to the density of sections with training data.

Multi-Class Reconstruction

The results of this experiment are found in Figure 4 and Column 2b of Table 3. The overall success score is 89.87%, which is mainly controlled by the predominant bedrock class (Class 6). Two other classes, esker and marine clay, demonstrate success rates over 50%. These units are somewhat better represented in the SVM training set than the remaining classes. Figure 5 shows that the success of reconstruction for a particular class in this experiment is almost directly proportional to the number of those class points in the training set, exceeding 75% when the number of training points exceeds 1% of the total.

We also compared area and volume of geological units in the original model and its reconstructed counterpart. The results presented in Table 4 and Figure 6 show that, for both area and volume calculations, the reconstructed and original values for any unit are the same order of magnitude. This suggests that along with single unit modeling, the SVM can efficiently be applied in multi-unit volumetric reconstructions.

Hyperparameter Sensitivity Tests and Multiple Parameter-Set Reconstructions

Figure 7 summarizes the results of grid search for the best pair of hyperparameters in binary and multi-class reconstructions. The best overall success rates, 97.79% and 92.03%, were achieved at [$C=2^1$, $\gamma=2^6$] and [$C=2^{-1}$, $\gamma=2^6$], respectively. The analysis of success rates in the class with the least number of training points yielded 77.38% and 22.46% at [$C=2^2$, $\gamma=2^6$] and [$C=2^9$, $\gamma=2^5$], correspondingly. As seen from the results, parameters are fairly stable, and a single range for C [2^{-3} - 2^{15}] and γ [2^4 -

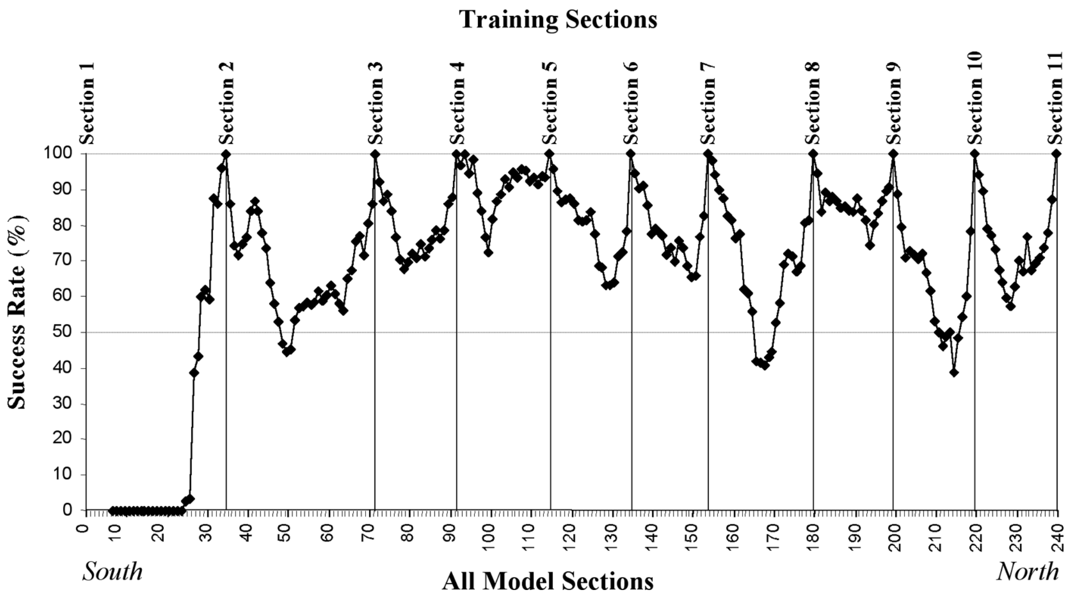


Figure 3. Validation results for binary esker reconstruction from 11 parallel sections. Success in Class 1 (Esker) against section number. Vertical lines indicate training sections. Total length of horizontal axis is 24km and sections are spaced at 100m.

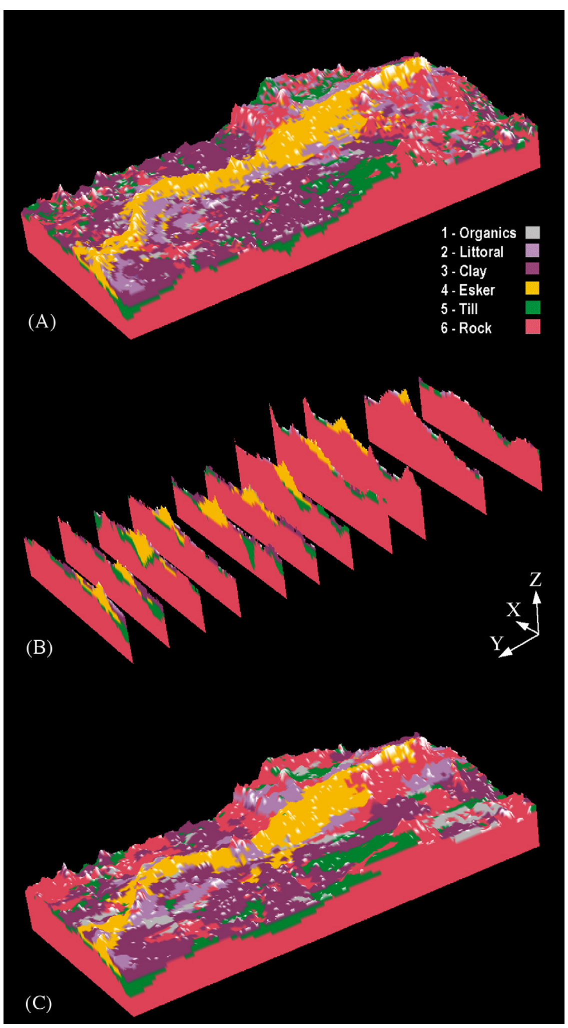


Figure 4. Multi-class esker reconstruction from 11 parallel sections. (A) Original model; (B) training set; (C) reconstructed model.

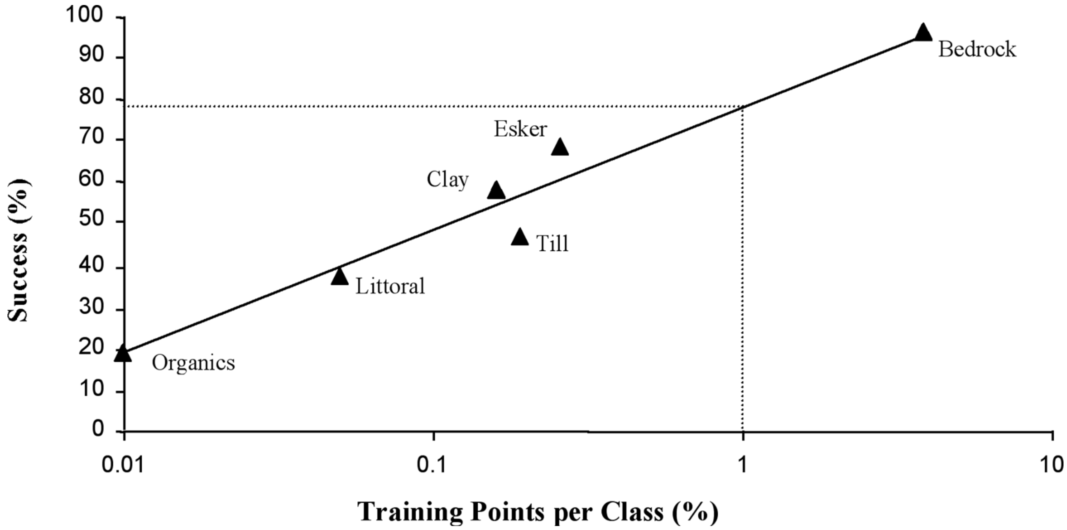


Figure 5. Validation results for multi-class esker reconstruction from 11 parallel sections. Success per class vs. number of training points per class.

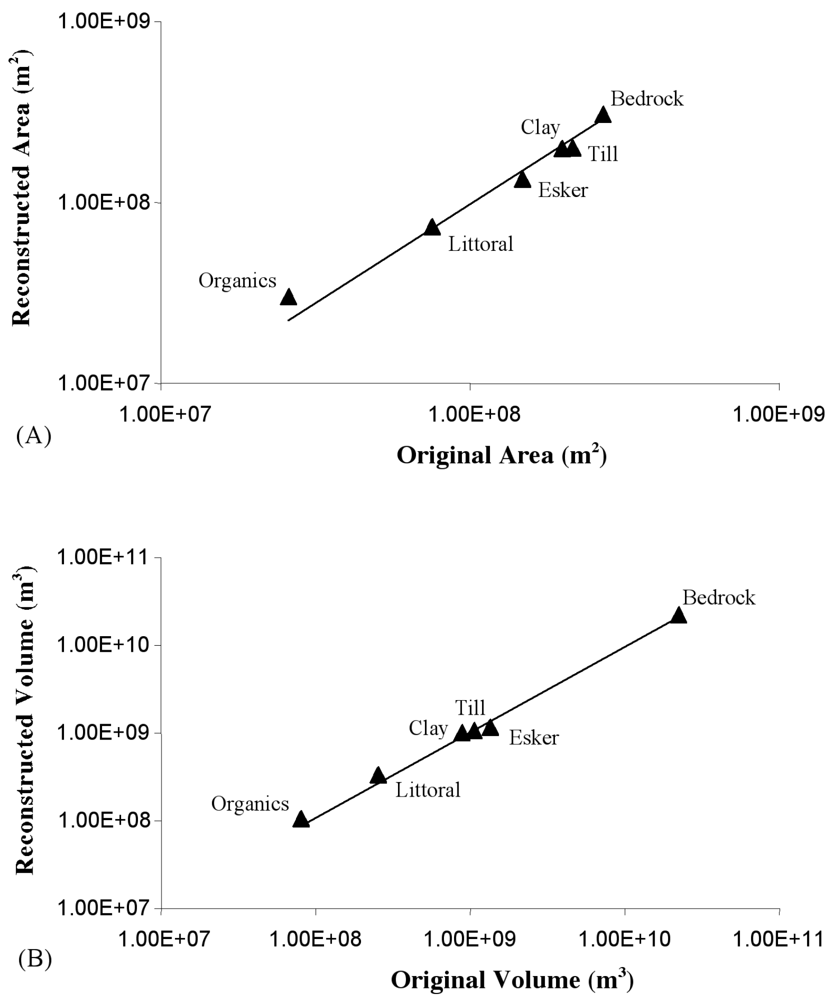


Figure 6. Surface area and volume comparison for original (reference model) and reconstructed geological units. (A) Surface area; (B) volume.

Table 4. Results of SVM reconstruction from 11 parallel sections for Esker/Abitibi. Unit area and volume comparison (original model vs. reconstructed). See Table 2 for training set statistics.

Geological Unit	1. Area (m^2)		2. Volume (m^3)	
	a. Original	b. Reconstructed	a. Original	b. Reconstructed
Organics	2.55E+07	2.99E+07	8.06E+07	1.05E+08
Littoral	7.56E+07	7.34E+07	2.54E+08	3.30E+08
Clay	1.99E+08	1.97E+08	8.86E+08	1.01E+09
Esker	1.48E+08	1.33E+08	1.35E+09	1.16E+09
Till	2.15E+08	1.99E+08	1.06E+09	1.07E+09
Bedrock	2.70E+08	3.06E+08	2.23E+10	2.23E+10

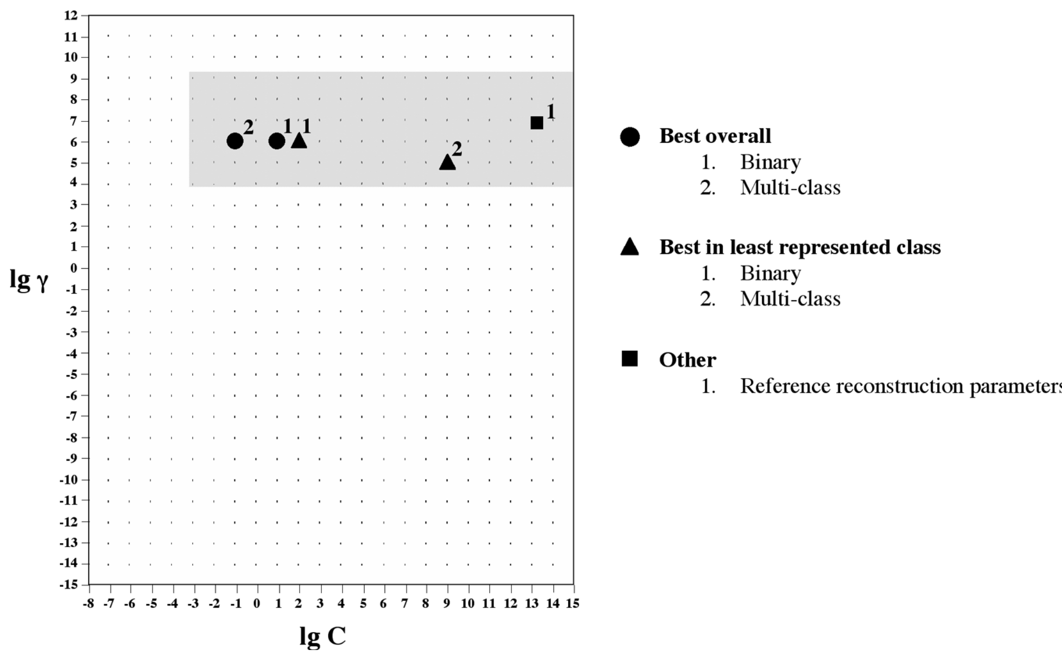


Figure 7. Summary of best results from parameter sensitivity tests for binary and multi-class reconstructions and proposed range for RBF kernel parameters (C [2^{-3}, 2^{15}] and γ [2^4, 2^9]).

2^9] can be recommended. Within this range, higher overall scores and higher scores for over-represented classes are achieved at somewhat lower C values. On the other hand, proper classification of points in the least represented classes requires higher C values. Visual examination of binary models built with combinations of parameters drawn from different corners of the above range also show that a more generalized picture can be achieved at lower C's (Figure 8a, 8b) while higher values of this parameter promote more detailed interpretation (Figure 8d, 8e). Average C values result in well-balanced models (Figure 8c). The influence of the second parameter (γ) is not as obvious.

CONCLUSIONS

Our experiments clearly showed that the SVM with RBF kernel can be efficiently used for both single- and multi-unit 3D reconstructions. The procedure is performed in a single step, which eliminates the need for unit-by-unit interpolation. Even from a limited training set (e.g., several cross-sections sparsely distributed across the study area) reasonable reconstruction results can be achieved.

It is important, however, that all classes to be reconstructed are reasonably represented in the training set. In the multi-class case, the reconstruction success was

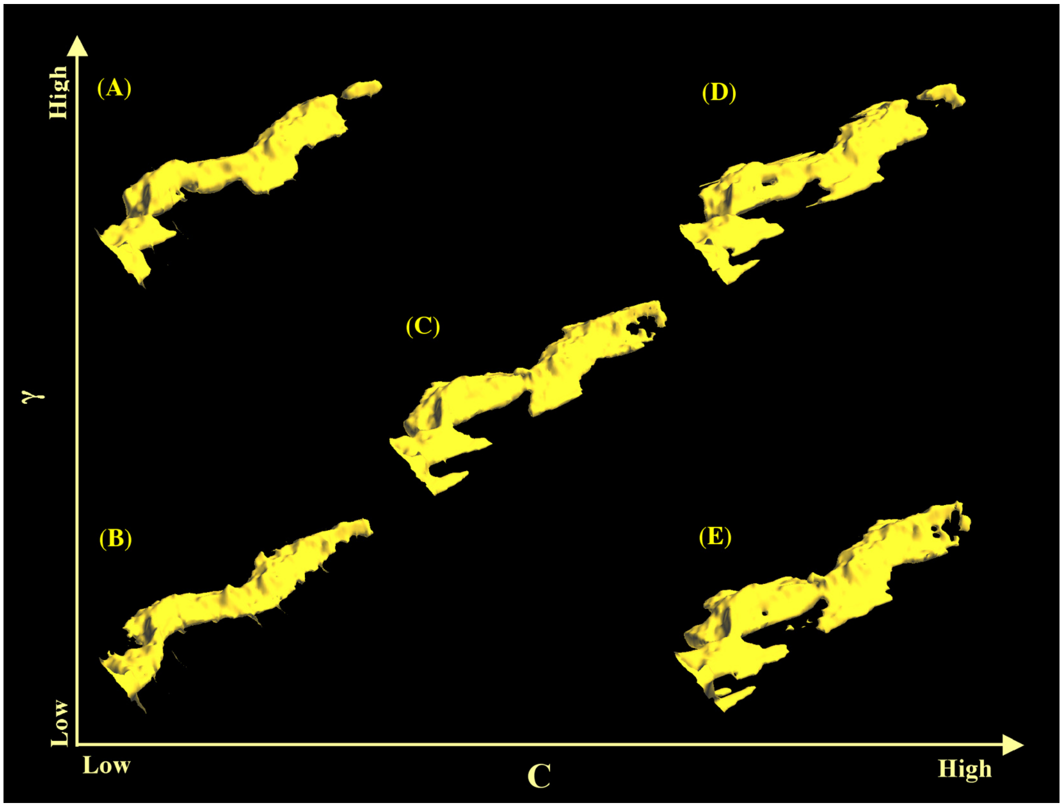


Figure 8. Binary reconstructions with parameters drawn from different corners of the range presented in Figure 7. (A) low C (2^{-2}) – high γ (2^{8}); (B) low C (2^{-2}) – low γ (2^{5}); (C) average C (2^{7}) – average γ (2^{6}); (D) high C (2^{14}) – high γ (2^{8}); (E) high C (2^{14}) – low γ (2^{5}).

shown to be directly proportional to the number of unit samples in the training data. The reliability of prediction is greater in the vicinity of the training data, and therefore, the density of training sections and spatial continuity of lithological units may directly affect the reconstruction results.

The kernel parameters should be chosen from the range 2^{-3}- 2^{15} for C and 2^{4}- 2^{9} for γ. When more model details are required or classes with a small number of training points are involved, higher C values should be considered. Lower C values result in more generalized models with fewer details. This favors classes that dominate the training set.

Finally, our results indicate that when appropriate parameters are chosen, not only the general shape of a geological body, but also such characteristics as its surface area and volume can be reconstructed with results close to those obtained from the application of classical GIS methods.

ACKNOWLEDGEMENTS

The research was conducted at the Quebec Division of the Geological Survey of Canada (GSC), and was funded by the Groundwater Program of the Earth Sciences Sector of Natural Resources Canada (contribution number 20060335). We are especially grateful to Andrée Bolduc (GSC) for her involvement in this work and for having made the Gocad® files with initial data and 3D Esker/Abitibi model used in experiments readily available.

REFERENCES

Abe, S., 2005, Support Vector Machines for Pattern Classification: Springer-Verlag, London, 343 p.

Bolduc A.M., Paradis S.J., Riverin M.N., Lefebvre R., and Michaud Y., 2005, A 3D esker geomodel for groundwater research: The case of the Saint-Mathieu–Berry esker, Abitibi, Quebec, Canada, in Russell, H., Berg, R.C., and Thorleifson, L.H. eds., Three-Dimensional Geological Mapping for Groundwater Applications, Geological Survey of Canada, Ottawa, Ontario, Open File 5048, p.17-20.

Chang, C-C., and Lin, C-J., 2001, LIBSVM: a Library for Support Vector Machines, accessed at http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf.

Cristianini, N., and Shawe-Taylor, J., 2000, Support Vector Machines: Cambridge University Press, 189 p.

El-Naqa, I., Yang, Y., Wernik, M.N., Galatsanos, N.P., and Nishikawa, R., 2002, Support vector machine learning for the detection of microcalcifications in mammograms: IEEE Transactions on Medical Imaging 21, p. 1552-1563.

Gilardi, N., Kanevski, M., Maignan, M., and Mayoraz, E., 1999, Environmental and Pollution Spatial Data Classification with Support Vector Machines and Geostatistics: Workshop W07 "Intellegent techniques for Spatio-Temporal Data Analysis in Environmental Applications", ACAI99, July, Greece, p. 43-51.

Hsu, C-W., Chang, C-C., and Lin, C-J., 2004, A Practical Guide to Support Vector Classification, accessed at http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

Hsu, C-W., and Lin, C-J., 2002, A comparison of methods for multi-class support vector machines: IEEE Transactions on Neural Networks 13(2), p. 415-425.

Mallet, J-L., 1989, Discrete Smooth Interpolation: ACM Transactions on Graphics 8(2), p. 121-144.

Noble, W.S, Kuehn, S., Thurman, R., Yu, M., and Stamatoyannopoulos, J., 2005, Predicting the in vivo signature of human gene regulatory sequences: Bioinformatics 21(1), p. 338-343.

Smirnoff, A., Boisvert, E., and Paradis, S.J., 2008, Support Vector Machine for 3D modelling from sparse geological information of various origins: Computers and Geosciences 34, p. 127-143.

Vapnik, V., 1995, The Nature of Statistical Learning Theory: Springer-Verlag, New York, 311 p.

Yu, X., Liong., S-Y., and Babovic, V., 2004, EC-SVM approach for real-time hydrologic forecasting: Journal of Hydroinformatics 06(3), p. 209-223.

Qualitative and Quantitative 3D Modeling of Surficial Materials at Multiple Scales

By Erik R. Venteris

Ohio Geological Survey
2045 Morse Rd., Bldg. C
Columbus, OH 43229
Telephone: (614) 265-6459
Fax: (614) 447-1918
e-mail: erik.venteris@dnr.state.oh.us

INTRODUCTION

An understanding of the distribution of surficial materials is key to many problems in geological engineering, mineral-resource inventory, and environmental remediation. However, for most states the zone between the surface (usually represented by surface glacial geology and the cooperative soil survey) and bedrock is unmapped. In Ohio, regional-scale maps of soil, surface glacial geology, and bedrock have been available for over 100 years, but concentrated mapping of the full sequence of surficial materials did not begin until 1998. The Ohio Division of Geological Survey (ODGS) is conducting three-dimensional (3D) mapping and modeling of surficial materials at the 1:100,000 scale. Currently, over one-third of the state has been mapped (concentrating on glaciated areas) at the 1:100,000 scale using qualitative methods based on geologic interpretation and drafting on Mylar (for an example of a completed product, see Swinford et al., this volume). Surficial materials are represented by two-dimensional (2D) polygons, which are assigned alphanumeric sequences describing sediment type, thickness, and lateral distribution ("stack" maps, Kempton, 1981), providing information in the third dimension. Mapping is conducted from the surface to the bedrock interface using soil maps, legacy geologic maps, water wells, bridge borings, and detailed site studies (mainly from environmental remediation). Envisioned applications for the GIS data and maps include surface/ground water simulations, mineral-resource inventory, and geologic engineering (seismic hazards, landslides, etc.).

As part of this mapping work, quantitative methods based on geostatistics are also being investigated. Lithology (clay, silt, sand, gravel) is modeled using sequential indicator simulation (Journel, 1983; Deutsch and Journel, 1998) to investigate methods for modeling stratigraphic and facies-scale variability at the 1:24,000-scale. Simulation is based on the same principles as kriging, but Monte Carlo techniques are used to develop multiple models (realizations) or configurations from one data set rather than obtaining a single, optimized estimate of lithology. Geostatistic simulation provides a range of statistically possible configurations of the subsurface that are faithful to the well data and statistical structure. While useful in themselves as 3D models, geostatistical models can give guidance to stack mapping by exposing consistent configurations of lithology. The amount of horizontal continuity in the models can give some measure of the uncertainty or appropriateness of assigning a stack sequence to potentially complex sediments of buried glacial valleys.

The goal of this paper is to illustrate the application of both methodologies being pursued at the ODGS to map unconsolidated sediments. Results are compared to enhance the understanding of the strengths and weaknesses of both approaches, and to discern how information from one technique can be used to improve the other. Geostatistical simulation has much to offer as a technique for 3D modeling. Output is in a 3D grid format ("voxel") and ready for volumetric (inventory) calculations or input into numerical models such as flow simulations for groundwater. There are scientific advantages as well; model parameters and procedures are completely traceable, so the maps are more "scientific" in that they meet the requirement of repeatability. While subjectivity is reduced, such modeling still requires much interpretation and "trade craft." A key advantage of simulation is the generation of multiple versions ("realizations") of the model that obey the data and the spatial structure. This aspect of simulation provides myriad possibilities for the assessment of uncertainty, ranging in complexity from generating basic statistics to describe variations between simulation runs to full analysis of the effect of uncertainty on all model parameters. Uncertainty assessment is key to evaluating risks when using the models for real world decisions. However, simulation techniques are not a way to generate maps quickly (good geostatistical practice requires careful, time-consuming investigation), nor map large areas (computation demands are limiting). Reconnaissance mapping over large areas is still the realm of traditional geologic mapping due to the limitations of the data (interpretation is needed over interpolation) and computation limits of large 3D voxel grids. This study proposes that both traditional mapping and geostatistical simulation have important and complementary roles in surficial mapping and characterization.

Mapping work for the 2006 fiscal year National Cooperative Geologic Mapping Program-STATEMAP component project is located in the Ashtabula and Youngstown 1:100,000 scale quadrangles located in northeast Ohio. This project area is the focus of this contribution. The area is heavily glaciated with extensive deposits of Wisconsinan- and Illinoian-age drift. Near Lake Erie, ice proximal (till, kames, outwash) and lake deposits (lacustrine, beach ridges) cover the Portage escarpment (Brockman, 1998). Further inland, the depositional environment changes to till plains and buried valleys. A key feature of most of the buried valleys in this region (Bagley, 1953) is that they were ice-dammed (to the north), which resulted in a greater portion of fine sediments (lacustrine deposits) than is found in buried valleys that drain to the south (Ritzi et al., 2000). A buried valley in the southwest corner of the 1:24,000 Ashtabula South quadrangle is the subject of both qualitative stack mapping and geostatistical modeling (Figure 1).

METHODOLOGY

A key component of both mapping techniques is the collection of base maps and boring data. A GIS/ digital database approach (based on ESRI ArcGIS and Microsoft Access) to data compilation is adopted for efficient distribution and storage. Detailed discussion of specific software modules and file formats used in the GIS data management is beyond the scope of this paper. Software used for geostatistical modeling is given a more thorough treatment.

The 1:100,000-scale “stack” maps (for this study, the USGS Ashtabula and Youngstown quadrangles) are initially drawn on 1:24,000-scale Mylar maps using (underlying) several different paper base maps and a light table. Interpretations are based on maps of soil parent materials, drift thickness, bedrock geology, and legacy geology maps. In addition, boring data from water wells (Ohio Division of Water, ODW), bridge borings (Ohio Department of Transportation, ODOT), and environmental studies (Ohio Environmental Protection Agency, OEPA) provided key information for mapping below the surface. The general data sources and procedures for making stack maps are described in the following sections.

Soils Maps

Whereas a major innovation of these maps is the inclusion of information in the subsurface, surface information remains critical and has a large impact on the appearance of the final map. Surface mapping units (lithologies) are largely derived from county-scale soil surveys. Soil survey information (1:15,840 scale) is used to make maps of parent material. Teams of pedologists use field investigations, soil sampling, and air photo interpretation to divide the landscape into polygons of like soils. The suite of soil mapping units and the rules for delineation on the landscape are based on a mutually agreed upon conceptual model. Mapping units are organized on major transitions in soil type, often those of significance to the management of the land. At the county scale, these transitions between soil units are usually due to changes in geomorphology and, therefore, provide a potentially high-resolution data source for surface lithology.

The primary source of digital soil data is the Natural Resource Conservation Service (NRCS) Soil Survey Geographic, commonly known as SSURGO (Soil Survey Staff, 2006). SSURGO GIS databases provide the mapping polygons as GIS files and extensive tabular data that describe soil horizonation, chemical/ physical properties, and descriptions of soil suitability for a wide range of land uses. The tables do not, however, explicitly define parent materials for each soil type. They are assigned to the polygons by creating a lookup table of interpreted (by this author) parent materials for each mapping unit. The tables are based on the detailed soil profiles and interpretive descriptions found in the written soil survey report. 1:24,000-scale maps of parent material are generated for each quadrangle (Figure 2). The stack model is much more accurate for layers near land surface than at depth because the level of detail in the soil survey is far greater than available well and boring data. However, the main intended use for soil mapping is land management. Therefore, there are often discrepancies between parent materials determined from soil polygons and the actual parent material (verification is conducted from boring data and by geomorphic interpretation). Parent material maps created in this fashion must be used with caution and interpreted with care.

Drift-Thickness Map

A second piece of mapping information is drift thickness (DT) (Powers and Swinford, 2004). DT maps are calculated from the surface digital elevation model (DEM) and the bedrock-topography map (Figure 3) (the bedrock topography map is usually updated and revised during the stack mapping process, past versions available as Mylar basemaps, contour shapefiles (vector), and grids (ODGS, 2003)). A key issue for DT maps is “flying outcrops” where the elevation of the bedrock topography (BT) exceeds that of the DEM. Such areas are not unexpected, as there are significant inaccuracies in both datasets. The accuracy of USGS DEMs is in the range of 5 to 10 feet Root Mean Squared Error (RMSE) (Venteris and Slater, 2005; Smith and Sandwell, 2003). The accuracy of BT within the Ashtabula and Trumbull quadrangles was estimated as

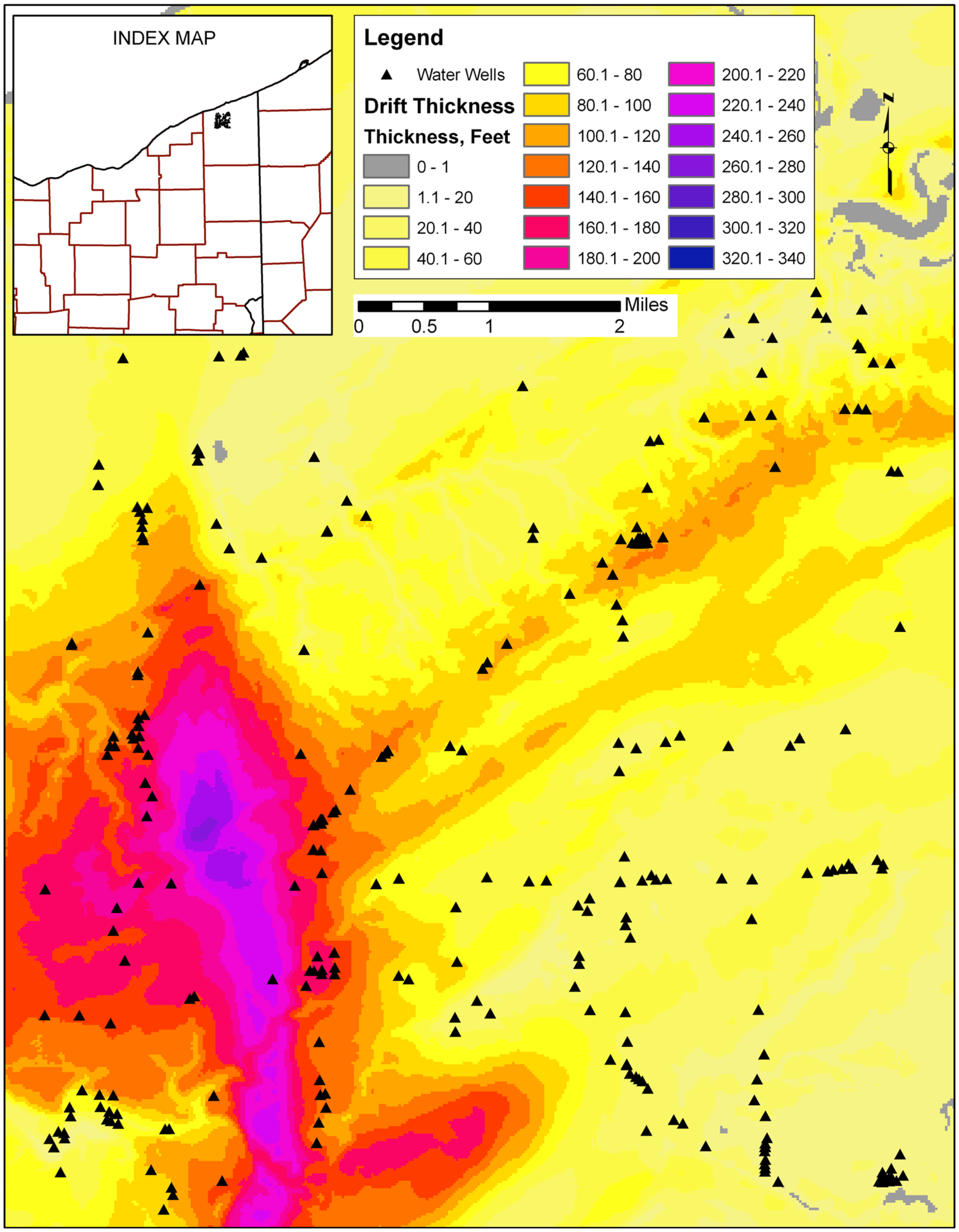


Figure 1. Map of drift thickness for part of the Ashtabula South 1:24,000-scale quadrangle, Ohio.

part of the current project. External validation was conducted for BT based on new bridge borings that were not used to create the map. Bedrock elevations from bridge borings were compared with interpolated elevations based on a TOPOGRID model of BT contours. RMSE error was found to be 22 feet. Assuming zero covariance between the error in the DEM and BT, the total error DT is

$$E^2_{DT} = \sigma^2_{DEM} + \sigma^2_{BT} \quad (1)$$

where σ^2_{DEM} is the error (expressed as a variance) in the DEM and σ^2_{BT} is the error in the BT. Assuming an error of 7 ft RMSE for the DEM, the total RMSE error in DT for this area is 23 feet. Hence, the error in DT is dominated by the error in the BT. Because of this uncertainty, drift thickness less than this value may in reality be areas of bedrock outcrop. For the current mapping project, areas of negative drift thickness are corrected to a DT value of zero feet. Areas of thin drift (less than 5 feet) are usu-

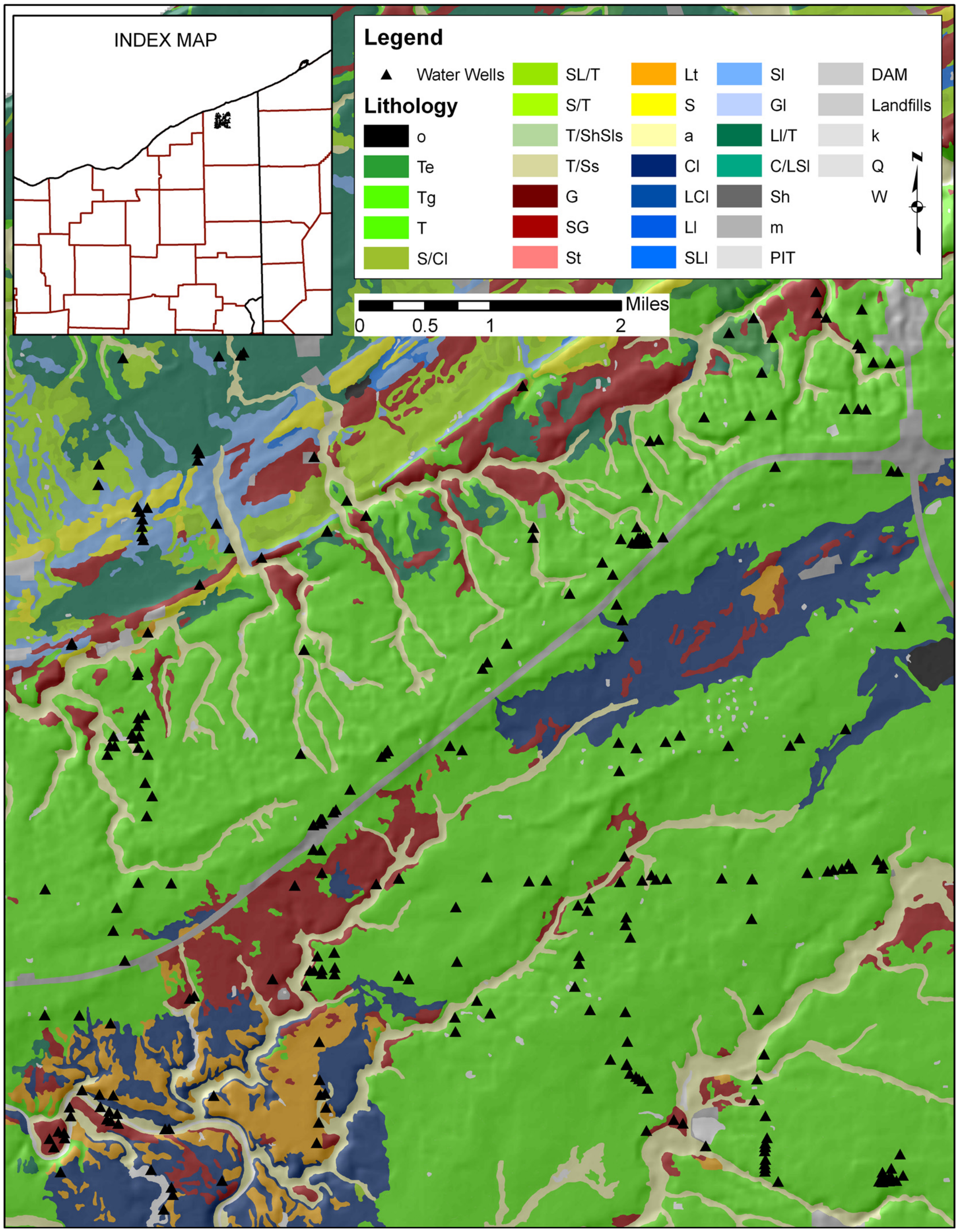


Figure 2. Parent material map based on SSURGO.

ally also identified in the parent-material (soil survey) maps, providing another means to confirm the presence of outcrop. Surficial materials at such locations are marked with parentheses on the "stack" maps to indicate that drift coverage is discontinuous in the area.

Other Base Maps

Legacy geologic maps are also used in mapping. Digital versions of the bedrock geology map (Ohio Division of Geological Survey, 2003) are used to map the bedrock base of each stack unit (the entire area of interest in the present mapping is underlain by shales of Devonian age). Existing maps of surface glacial deposits are compared to the parent material maps (from the soil survey) to aid in the assignment of surface units, especially in interpreting depositional environment. Often, lithology information is obtained from the well data and the soils maps without a

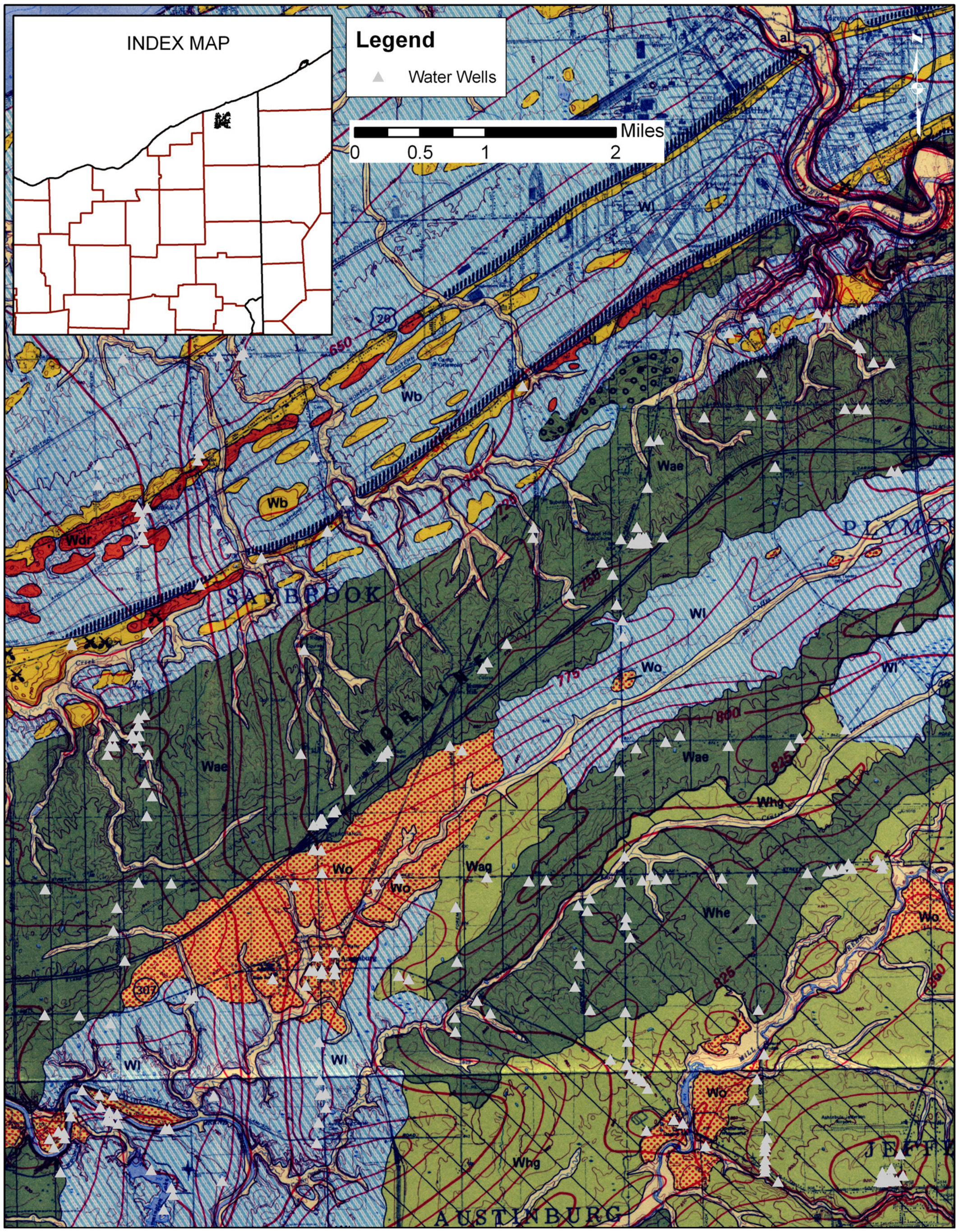


Figure 3. Glacial-geologic map of the study area taken from White and Totten (1979).

geomorphic interpretation. However, such interpretations are critical to the use of this data for its intended applications. For example, identification of sand and gravel (a lithology) as either an outwash or ice contact deposit (a geomorphic interpretation) is key to groundwater modeling, as there will be marked differences in facies hydraulic structure and conductivity between the materials.

County-scale glacial-geology maps (Figure 3) were available for the three counties of the 2006 STATEMAP project area (White, 1971; White and Totten, 1979; Totten and White, 1987). These maps were scanned and rectified (not digitized in a vector format) and used in the GIS as an additional layer to aid in the mapping of surface units (assignment of lithology, etc).

Water-Well and Other Boring Data

Mapping in the subsurface is based on lithology and other information available from borings drilled for water wells and engineering studies. The most spatially dense data set are water well records from Ohio Division of Water (ODOW). It is a legal requirement that the records (logs) from the drilling of water wells be filed with the state. The lithology (texture), thickness, and occasionally color of layers encountered while drilling the well are contained in the records. The records are obtained as Excel spreadsheets (ODOW, 2005) and converted to a geodatabase for use in ArcGIS. An ArcGIS Visual Basic application was built to display the well location and lithology on paper base maps (Figure 4). The water well data provide critical information on lithologic sequences with depth, but the unit "clay" requires careful interpretation. It is likely that the "clay" unit of the water-well records contains lithologies that range from clay to silt. The identification of silt units is strongly underrepresented compared to the proportion indicated in more detailed and reliable texture data (textures based on laboratory work) such as those from bridge borings (Table 1). The lithology "clay," therefore, is reinterpreted in this modeling to mean clay and silt.

More accurate and detailed depth information is available from the Ohio Department of Transportation in the form of detailed records from geotechnical borings drilled to support bridge construction. These data are available as paper records from ODOT (Figure 5). A digital relational database (Figure 6) is in development

Table 1. Comparison between sediment textures in water-well and bridge-borings in Ashtabula County. The comparison contains many sources of serious bias, as the water-well data generally extends to greater depths than the ODOT bridge boring data, and the spatial distribution of bridge borings is seriously biased. However, it is clear the silt is grossly underrepresented in water wells. The column "Bridge Boring" gives the textural percentage used to define each lithlogic class in the bridge borings.

Texture Class	Water Well	Bridge Boring	Cutoff for Bridge Boring
Clay	0.6	0.37	>40%
Silt	0.004	0.26	>40%
Sand	0.18	0.07	>40%
Gravel	0.21	0.07	>30% sand, >10% gravel

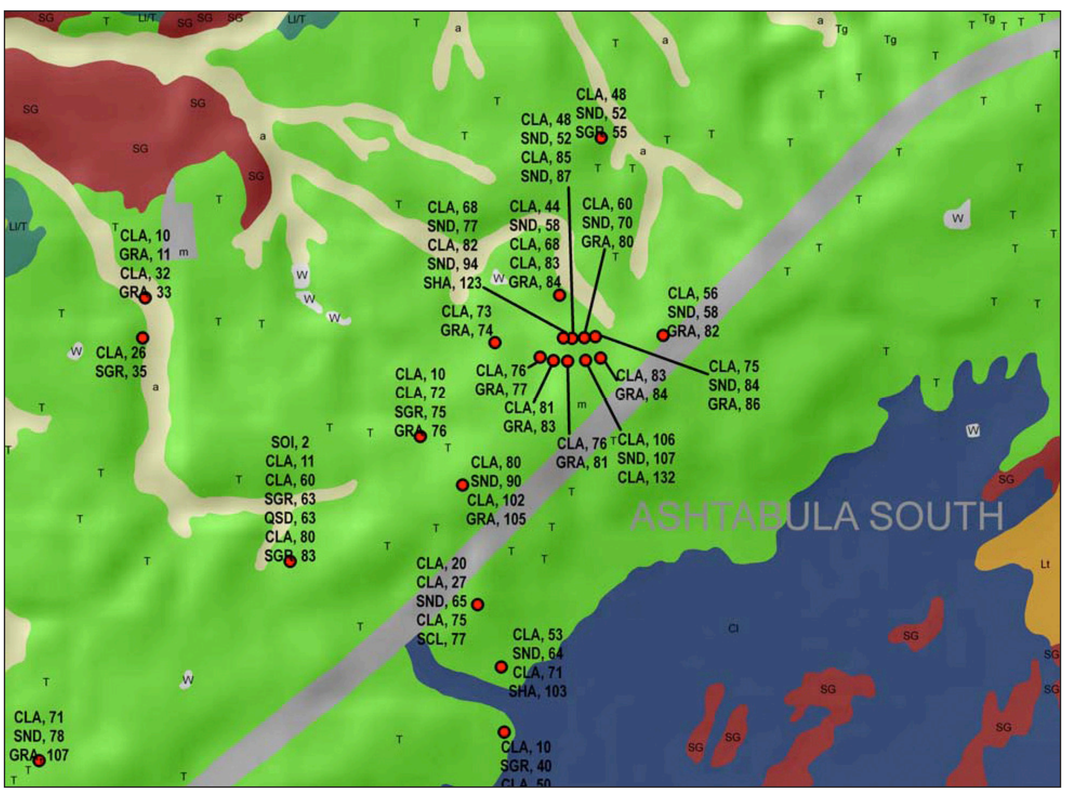
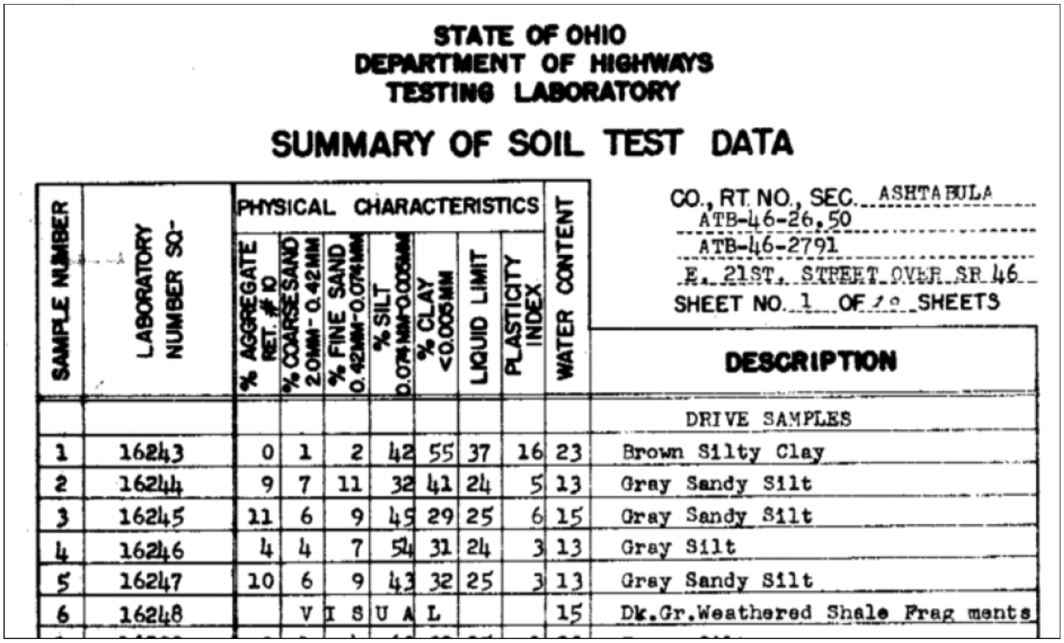


Figure 4. Close-up example of water-well postings.



STATE OF OHIO
DEPARTMENT OF HIGHWAYS
TESTING LABORATORY

SUMMARY OF SOIL TEST DATA

CO., RT. NO., SEC. ASHTABULA
ATB-46-26.50
ATB-46-2791
E. 21ST. STREET OVER SR 46
SHEET NO. 1 OF 10 SHEETS

SAMPLE NUMBER	LABORATORY NUMBER SQ-	% AGGREGATE RET. # 10	% COARSE SAND 2.0MM - 0.42MM	% FINE SAND 0.42MM-0.074MM	% SILT 0.074MM-0.005MM	% CLAY <0.005MM	LIQUID LIMIT	PLASTICITY INDEX	WATER CONTENT	DESCRIPTION
										DRIVE SAMPLES
1	16243	0	1	2	42	55	37	16	23	Brown Silty Clay
2	16244	9	7	11	32	41	24	5	13	Gray Sandy Silt
3	16245	11	6	9	49	29	25	6	15	Gray Sandy Silt
4	16246	4	4	7	54	31	24	3	13	Gray Silt
5	16247	10	6	9	43	32	25	3	13	Gray Sandy Silt
6	16248		V	I S	U	A L			15	Dk.Gr.Weathered Shale Frag ments

Figure 5. Scan of a portion of an ODOT bridge boring record.

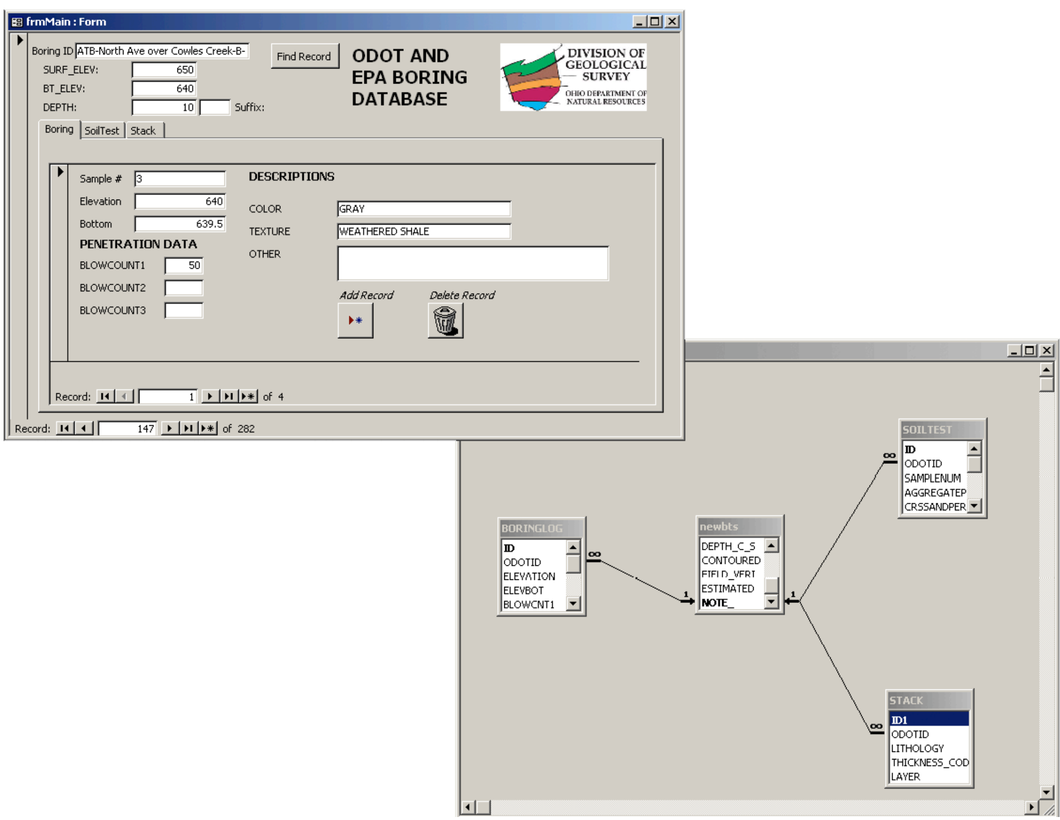


Figure 6. Schematic showing fields, tables and database relationships for the bridge boring/ OEPA database, built by the Ohio Division of Geological Survey.

by ODGS to facilitate computer-based analysis of these data. ODOT records provide detailed engineering data on texture, mechanical strength (blow counts, plasticity indexes, liquid limit, wetness), and lithology (color, texture, texture class). ODOT bridge borings provide excellent site information, but have limited spatial density and a strong locational bias. The borings are mainly collected where roads intersect major streams and the depth is usually limited to approximately 50 ft, which is insufficient in typical buried valleys where drift thickness can exceed 300 feet.

Further depth information is available from OEPA in the form of detailed site studies. Engineering firms with projects such as sanitary landfills and industrial facilities file with OEPA their detailed studies of the subsurface. These reports typically contain several well borings (typically 10 to 30 wells) with a wide range of engineering data. The types of information are usually similar to ODOT records, but with inconsistent coverage (one well may contain good textural information, but lack blow counts, etc.). Digital capture of all wells in these sites is beyond the mapping goals (site studies vs. regional mapping) and resources (time/ labor constraints) of the ODGS. Typically, the best well of the group (with represtative geology and data compatible with the ODOT database) is chosen from each site and entered into the same database as the bridge borings.

Qualitative Mapping—Ashtabula South Quadrangle Case Study

The idea for mapping in three dimensions using stack sequences and 2D polygons is based on previous surficial mapping approaches (Kempton, 1981), with refinements unique to ODGS (Brockman et al., 2004). The emphasis in ODGS work is on lithologic characterization, so stack sequences describe both layer lithology and thickness (estimated to within 50%) from the surface to bedrock. Little attempt is made in this mapping work to assign time units to the layers, except where distinctions between the Wisconsinan and prior glaciations are well known or obvious. Tills are not mapped by traditional stratigraphic units (for example, the Hiram and Waverly tills of White and Totten, 1979) but are divided where there are significant changes in texture or chemistry (carbonate content).

Base data are compiled into a stack model using traditional geologic mapping methods (drafting) followed by GIS digitization. Experience has shown that accurate and efficient generalization and interpretation require the geologist to utilize much information in a spatial context. Large scale (1:24:000) mapping conducted using transparent Mylar and paper base maps can display much more information at a legible scale at one time than any practical (inexpensive) computer screen. After compilation, hand-drawn maps are digitized and attributed to make GIS coverages using standard GIS techniques.

The first step in mapping is to delineate the major lithologic and geomorphic units present at the surface. Polygons of surface features are drawn initially by generalizing (as appropriate for a 1:100,000-scale map) the parent material polygons from the interpreted soil-survey map. Elevation contours and DEMs are often used as an additional guide to generalization (breaks in slope, or stream and erosion patterns). The surface model is further refined by adding geomorphic interpretations based on the mapper’s own knowledge, aided by legacy geologic maps such as the glacial-geology series. The surface model (Figure 7) is also checked and verified using information from the various water-well logs and other boring data.

The difficulty and need for geologic interpretation increases greatly when mapping in the subsurface. Subsurface transitions are delineated on the maps using a contrasting line color, which is expressed as a different line style on the final map (solid lines for changes in surface materials and dashed lines for subsurface transitions (Brockman et al., 2004; Swinford et al., this volume)). Subsurface polygons denote large changes in thickness and lithologic sequence. The first step in subsurface mapping is inspection of the drift thickness map. The geologist looks for major geomorphic features, which provide a rough idea of where the major transitions will be drawn. In general, breaks in thickness that delineate buried valleys and end moraines are the most common and critical to communicating the geology of the area. Once major thickness transitions are denoted, a stratigraphic model is developed to assign stack sequences to each mapping polygon. This model is based on inspection and analysis of the well data in the area. In general, detailed information (mainly texture, but penetration and plasticity tests are also useful) from bridge borings and environmental study sites are used to develop an initial conceptual model. The model is then verified and extended spatially using the more general water well data (which are usually posted to the parent material and drift thickness base maps) and an understanding of the expected configuration of sediments for the given geomorphologic environment.

An illustrative example of mapping at depth is provided for the SW corner of the Ashtabula South Quadrangle. This area is unusually complex, as it contains the Painesville end moraine superimposed on the buried valley associated with the Grand River. The area surrounding these features contains end moraines, beach ridges, and lacustrine sediments. The main subsurface feature, a major north-south bedrock valley, was dammed (to the north) when glacial ice occupied the Lake Erie basin. The feature is approximately four miles wide and contains drift with a thickness up to 300 feet (Figure 1). When mapping this buried valley, we have to make two major decisions:

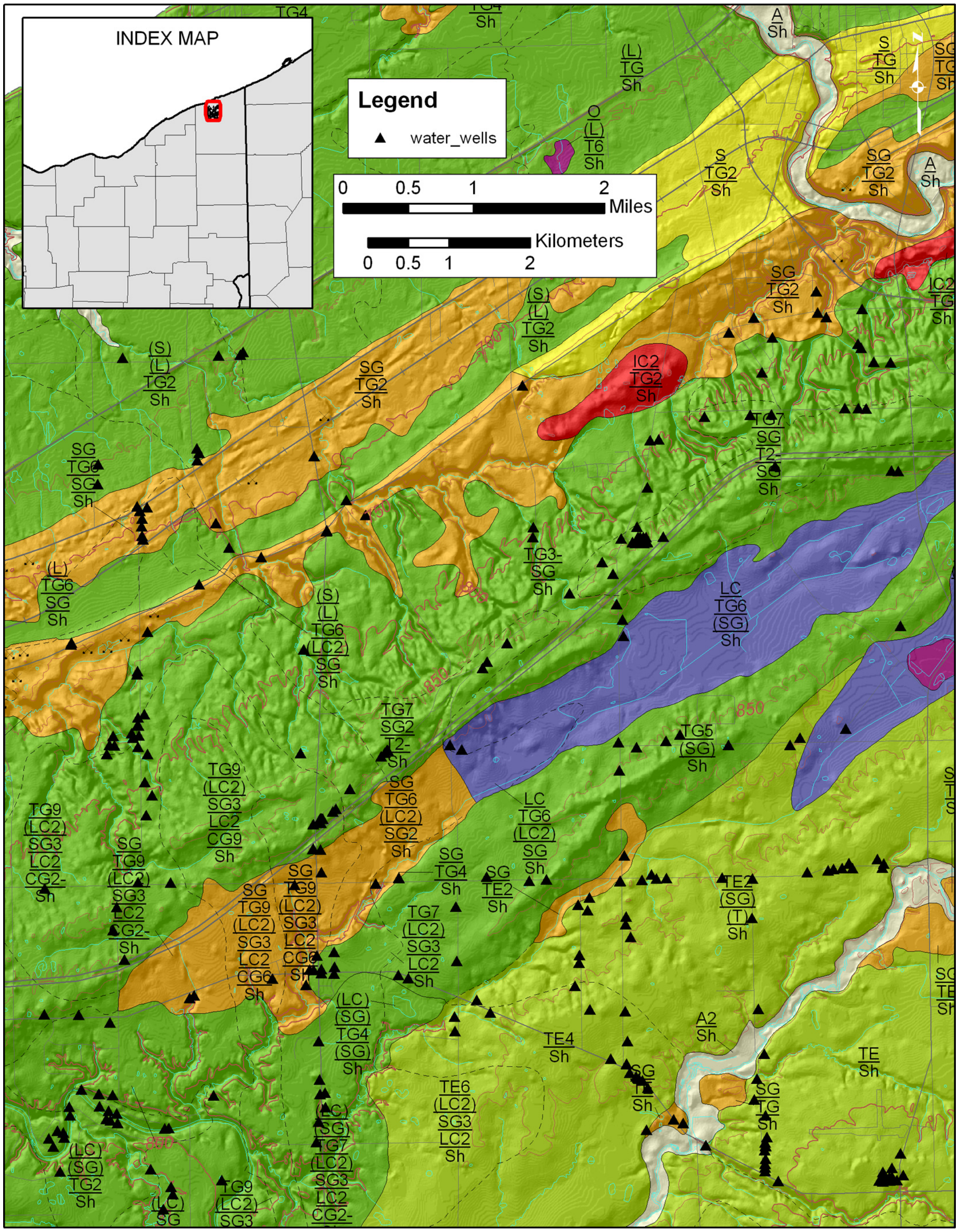


Figure 7. Preliminary model of the subsurface geology at head of the Grand River. This is a portion of the “stack” map for the 1:100,000 scale Ashtabula Quadrangle. The map is currently under review. Some key abbreviations: TG = Wisconsinan till unit high in silt content, TE = Wisconsinan till unit high in clay, SG = sand and gravel, LC = silt and clay (generally lacustrine), Sh = shale bedrock, S = Sand, CG = buried-valley deposit with undifferentiated lithology.

1. What is the best way to draw polygons to communicate the important transitions in thickness and lithology?
2. How do we best generalize available well data into a stratigraphic model to assign stack sequences to these polygons?

Maps of drift thickness are used to provide an initial impression of the subsurface configuration. The DT map (Figure 1) shows the superposition of the north-trending Grand River buried valley and the SW-NE trending Painesville moraine. Some "fingers" of increasing drift thickness off the main valley are due to the Painesville moraine, whereas others are due to the presence of buried side valleys. The bedrock-topography map must be used in conjunction with the drift thickness map for proper interpretation. These maps were used to develop basic ideas on mapping the subsurface, such as the locations and extents of major subsurface polygons, total depth needed for stack sequences, boundaries of major subsurface morphologic units (in this case delineating a reasonable extent of lake sediments when the buried valley was flooded), and cartographic concerns such as minimizing small and sliver polygons caused by the interaction of surface features and subsurface polygons.

Borehole information is used to develop a generalized stratigraphic sequence or "stack" for each polygon. Each stack contains the lithology (and geomorphology, where applicable) and thickness estimate (within ±50%). The goal is to generalize information from many wells into vertical sequences and identify horizontal transitions and breaks in the sequences (overall depth or lithology) that warrant drawing additional polygon boundaries.

Stack sequences for the study area were mainly based on water-well lithology logs. Detailed (bridge) boring data were not available for this geomorphic feature within the quadrangle boundary. Only one detailed site with limited depth (~100 ft) existed to the southwest (in the Jeffersonville 1:24,000-scale quadrangle), so stratigraphic models were developed mainly from the water-well database. The first attempt at a stratigraphic model was based on interpretation of common patterns in lithology. An attempt to correlate first clays, first gravels, second clays, second gravels and so on was done. Some patterns emerged, but it was clear there was much variability between the data. Differences in vertical resolution and quality between loggers, complex spatial variations in geology, and blunders all made it difficult to create a generalized stack. The preliminary subsurface model and stratigraphic sequence (Figure 7) was subject to a more rigorous review, aided by results of geostatistical and statistical analysis.

Quantitative Mapping—Geostatistical Modeling (Sequential Indicator Simulation)

Geostatistical simulation techniques exist to model and simulate categorical variables such as lithology. Such methods are common in oil and gas exploration (Deutsch, 2000) and have been used to characterize surficial deposits for ground-water modeling (Carle and Fogg, 1996; Ritzi et al., 2000). The various forms of geostatistical simulation are generally preferred over indicator kriging for modeling of surficial deposits because buried-valley and other surficial deposits have high spatial complexity. In addition, the sample spacings of typical well data sets contain gaps or average spacings that greatly exceed the scale of autocorrelation. In this context, the rigorous techniques of uncertainty analysis and superior ability to extrapolate results beyond the well data (due to sequential approach used in simulation algorithms) of simulation techniques are valuable. While model results in sparsely sampled regions are not reliable for predicting the position of lithologies (say for drill planning), such extrapolation is useful for groundwater simulation, provided many realizations are used to understand the range of possible results in poorly constrained areas. There are also many theoretical reasons (missing variance and inherent smoothing of kriging, etc.) to choose simulation techniques over kriging (Deutsch, 2000).

The goals and output results of geostatisical simulation are different than those of indicator kriging. Kriging provides the best estimate of values at unsampled locations. The goals of simulation are to use kriging in conjunction with Monte Carlo techniques to produce many different realizations faithful to the data locations and reproduce global statistics (histogram) and local spatial structure (variogram). Each realization represents a statistically valid, potential configuration of the subsurface. However, interpolated values in each model are not optimal estimates and can vary widely between realizations. The variation between realizations is the strength of the method, as it is the basis for evaluating the uncertainty of the model. The more tightly constrained the model (large amount of well control, predictable spatial structure with strong (repeated) patterns), the less variability between runs. Summary statistics that characterize the differences between realizations provide a rigorous and convenient way to assess model uncertainty.

A full development of sequential indicator simulation (SISIM) (Journel, 1983) is not presented here. However, the basics of the algorithm are described to provide the reader with insight as to how the technique works and the configuration of the data and spatial autocorrelation will affect results. First, a regular three-dimensional grid is specified for the volume of interest (for efficiency) and a random interpolation order is chosen for the cells. This is important, as simulated values are treated as data points and used to calculate kriging weights for subsequent cells. SISIM is based on the calculation of a conditional distribution for each cell, which is randomly drawn from to assign a lithology. The conditional distribution is based on the kriging estimate (local information) and the global probability. For each lithological type (***k***) at location (**u**), the conditional probability is

$$p_k^*(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha i(\mathbf{u}_\alpha ; k) + \left(1 - \sum_{\alpha=1}^{n} \lambda_\alpha\right) p_k(\mathbf{u}) \qquad (2)$$

where λ_α is the simple-kriging weight, i is the indicator code (a binary variable for each lithology, i.e., sand (1) or not sand (0)) for a neighboring data point (either a "real" data point, or a simulated cell) and $p_k(\mathbf{u})$ is the global probability for the respective lithology. Hence, the conditional probability is a function of the values of neighboring cells (whose influence with distance is defined by the variogram and kriging) and a global probability. Next, p^*_k results for each lithology are combined to define the cumulative conditional distribution function (ccdf). A random number between 0 and 1 is drawn, and the ccdf is used to assign a lithology for that cell.

Some key aspects of the algorithm should be considered when evaluating an individual realization:

1. Global probabilities have decreasing influence with increasing density of neighboring data (due to the presence of well control or cells that are filled in later in the grid order). In data-rich regions, results are mainly influenced by the data values and the variogram (the distance of influence increases with range; the nugget effect (if used) decreases the kriging weights and increases the influence of global proportions). In sparsely sampled areas, initial cells are controlled by the global probabilities.
2. Randomness and subsequent differences between realizations arises from two sources: the random path for assigning cell values and the random draw to assign a lithology from the conditional distribution function. Once a lithology is assigned to an empty cell, it becomes a data point and influences the neighboring results. This is generally a good feature of sequential simulation, as it tends to produce geologic bodies even in areas poorly constrained by wells (for example, once sand is assigned to a cell, it is more likely that neighboring cells also will be assigned as sand, creating a sand layer). The end result is that, even in areas of little or no data, global statistics and spatial structure are preserved.

The sequential indicator simulation algorithm in GSLIB (Deutsch and Journel, 1998) was used to model lithology (clay, silt, sand, or gravel) for the southwest portion of the Ashtabula South Quadrangle described above.

Workflow

The work presented here was a preliminary exploration, mainly intended to create 3D models for comparison with the stack maps. A far more rigorous modeling of the area (conducted in elevation space, using simulations nested by stratigraphic units, and models of spatial structure based on well statistics (Ritzi et. al, 2000)) is presented in Venteris, 2007. The preliminary investigations presented here were conducted to test whether lithologic information from the water wells could be extended (interpolated) into continuous models using geostatistical simulation techniques. The amount of similarity between models created by the two techniques could provide insight into the geology of the study area and the quality of information contained in water wells.

The first step in modeling was to define the spatial domain of the model. The area for study was chosen for geological interest and the presence of sufficient water-well data to support simulation. The lack of detailed ODOT borings was a significant disadvantage to this study area. The models were created in depth-space rather than using real world elevations. Such an approach was advantageous for preserving lateral continuity for layers that follow topography such as till sheets and for comparison with the stack model, which were also modeled as depth and thickness. A three dimensional grid with cell dimensions [x=200 ft, y=200 ft, z=2 ft] (data continuity is much higher in the z direction) was defined for modeling. This domain contained both unconsolidated sediments and bedrock, the boundary between the two defined by the DT grid. A FORTRAN program was written to convert the 2D ASCII format DT grid into a 3D GSLIB format voxel, which was then used to clip the model results. Geostatistical simulation was only conducted for the portion of the data containing unconsolidated sediments. Bedrock and unconsolidated sediments were not modeled together because the probability of either lithology group is not spatially constant over the domain (stationarity).

Well data were converted to a data format appropriate for geostatistical modeling. Water well data from ODOW were converted to a simple set of indicator codes. The conversion required some interpretation and generalization. For example, a water well record described as C/R or "Clay and Rock" was coded as clay, assuming that the driller was describing a till with rock fragments. There were many lithologies that require interpretation in the water-well database, but their overall proportion in the dataset was small. Most descriptions were of common and easily classified lithologies. Unconsolidated lithology classes from the water wells were reclassified into four indicator variables [clay=1, silt=2, sand=3, gravel=4]. In addition, the wells were discretized in the vertical direction. Lithologies for wells in the ODOW database were given a range of depth (upper and lower values). The water wells were discretized at one-foot increments to provide sufficient continuity of lithology values for the intended vertical resolution of the voxel model.

SISIM required the assignment of global probabilities for each lithology. For this data set, clay was the dominant lithology (Table 2). Silt is grossly under represented in the water-well dataset as discussed above, and so cells modeled as clay include both clay and silt. The data were checked for clustering bias using the DECLUS routine in GSLIB. Bias due to clustering is less than 5% for this dataset and is not considered a significant source of error.

The next step was to define the spatial structure (variogram) for use in assigning kriging weights. First, experimental variograms were calculated from the data. Experimental variograms were used to create model variograms for input into the simulation (kriging) procedure. Model variograms can be created by visual estimation, trial and error modeling, and automatic fitting routines. Key information to obtain from the experimental variograms was the overall shape (expressed as a function, usually spherical), the range of autocorrelation (where the variogram intercepts the sill), and the magnitude of the nugget effect (non-zero intercept, caused by small-scale variability below the distance of the lag spacing and measurement error).

Experimental variograms were calculated (using GAMV routine in GSLIB) for each lithology (except silt, for which there are insufficient data points) in the vertical and horizontal directions. A range of experimental variograms were calculated to explore many possible scales of spatial structure (by adjusting lag spacings and the number of lags) and to check for anisotropy. There was some indication of anisotropy (semi-variance values exceeding sill), but the noise in the data set precluded an accurate estimate of directionality. Indicator variograms in the vertical direction (Figure 8) were generally smooth,

Table 2. Proportions of each lithologic unit for the simulated area.

Texture Class	Water Well
Clay	0.816
Silt	0.002
Sand	0.137
Gravel	0.045

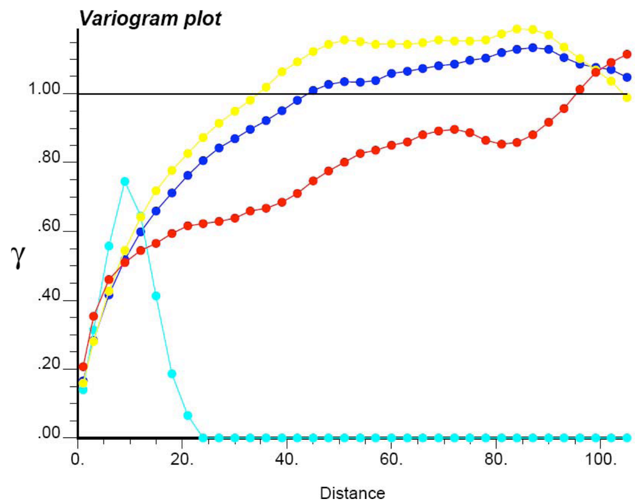


Figure 8. Indicator variograms in the vertical direction. Dark blue is clay, light blue is silt, yellow is sand, and red is gravel.

had a small nugget effect, and were fit with a basic spherical model (Table 3). Experimental variograms were much less stable in the horizontal direction. Smooth variograms were possible using large lag spacings (Figure 9) but a more revealing picture was obtained by using small lag spacings (Figure 10). The smooth experimental variograms suggested a very large nugget effect and a range of around 3,000 feet. The gravel lithology never fully approached the sill, which suggested an anisotropic structure. Experimental variograms using a smaller lag spacing showed a more complex picture. For small lag spacings the nugget was much reduced, but the semi-variance oscillated widely, making range selection ambiguous. This "hole effect" could have been due to the natural spatial structure of the glacial sediments or a result of incomplete and noisy sample data. Determining the range (point of intercept with the sill) was essential for kriging and simulation. The variograms showed an initial structure (local maxima) at about 1,500 feet. All variograms intercepted the sill several times over the range of 1,500 to 5,000 feet.

Several variograms were used in the simulations, as the experimental variography did not provide clear guidance for the horizontal variogram model (a different, more rigorous approach to characterizing spatial structure for SISIM modeling is presented in Venteris, 2007). Three example variogram models were used to demonstrate a reasonable range of results. No attempt was made to choose the best model from the range of possibilities. Rather, multiple scenarios were run to illustrate the effect of variogram parameters on results. Model variograms for simulation were loosely based on the experimental horizontal variograms (vertical models are held constant (Table 3)). For the first experiment, the horizontal range for all lithologies was set to 2,000 feet (a compromise

Table 3. Model variogram parameters used in the three indicator simulation runs. The nugget effect is zero for the short range and long range models. The nugget is 0.5 for the short-range with nugget model.

Indicator	Horiz. Range (Anisotropic)	Vertical Range	Contribution
Nugget			
1	2000	50	0.5
2	2000	15	0.5
3	2000	40	0.5
4	2000	80	0.5
Short Range			
1	2000	50	1
2	2000	15	1
3	2000	40	1
4	2000	80	1
Long Range			
1	5000	50	1
2	5000	15	1
3	5000	40	1
4	5000	80	1

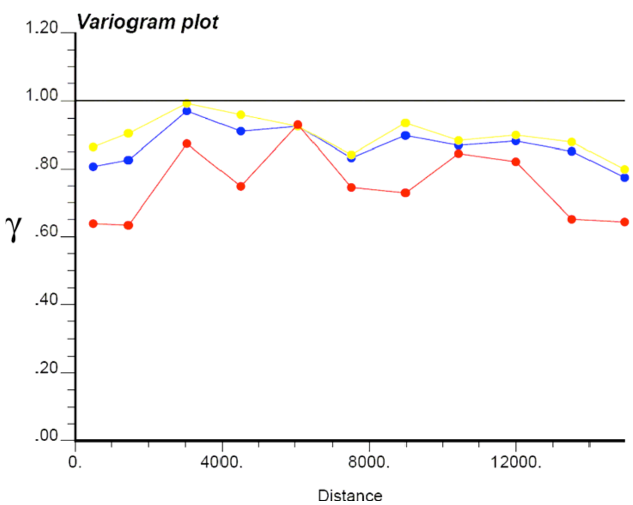


Figure 9. Indicator variograms in the horizontal direction (isotropic) using a lag separation of 1500 feet and a lag tolerance of 800 feet. Colors and lithologies are the same as in Figure 8.

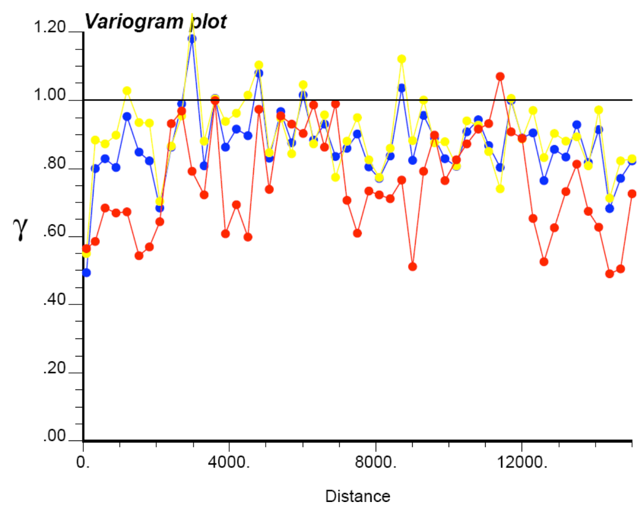


Figure 10. Indicator variograms in the horizontal direction (isotropic) using a lag separation of 300 feet and a lag tolerance of 150 feet. Colors and lithologies are the same as in Figure 8.

between the ranges of clay and sand (intercept at 1,500 feet) and gravel (around 2,500 feet)) with a large nugget contribution (0.5, or 50% of the variance due to measurement error and spatial variation below the scale of the lag distance). For the second model, the range was held to 2,000 feet, but the nugget effect was set to zero (assuming nugget due to inadequate sampling rather than geologic variability). For the final model, it was assumed that both early oscillations and large nugget effects were spurious. The variograms were modeled with a range of 5,000 feet and a nugget effect of zero. This model represented the maximum amount of spatial continuity that could reasonably be interpreted from the experimental variogram results.

Lithologies were simulated using the SISIM algorithm as implemented in GSLIB. 16 individual realizations were produced for each variogram model. As a rule of thumb, the number of realizations should be around

100. Running this model in GSLIB resulted in a stack overflow and program discontinuation after 16 runs. The solution was to run multiple batches and then write software to recombine them for final products. Modified software to process 100 runs was not complete at the writing of this contribution. The goals of this study are mainly exploratory and illustrative, so the limitation to 16 realizations had little meaningful impact.

To find the most common value and assess the variability of the realizations for each cell, the realizations were post-processed. Firstly, the portion of the model at or below bedrock was removed (clipped). FORTRAN programs were written to post-process the SISIM runs. The mode value was used as the most common value between runs. The algorithm did not break ties between lithologies (a very rare occurrence), and cells with an ambiguous mode were written to "no value." Cells that did not contain a well data point could take any value from realization to realization. The variety of lithologies written to each cell was of interest. Variation between runs for each voxel cell was evaluated using Shannon's (1948) entropy where p_i was the proportion of each lithology within the

$$H = -\sum_{i=1}^{n} p_i \ln p_i \qquad (3)$$

set of realizations. Values close to zero indicated consistent values between runs, and values near 1 represented high variation between runs.

RESULTS

Qualitative Mapping

A wide range of surface features of mappable size exist in the study area. The preliminary stack map is presented in Figure 7. The major surface geomorphic feature is the Painesville end moraine, which trends from the southeast to the northwest. On the lake ward side (to the northwest), sand and gravel beach ridges are superimposed. Most alluvial deposits are too small to be mapped at 1:100,000 scale. Behind the end moraine is a small southeast/ northwest trending lacustrine deposit. In the far southwest corner is a large sand and gravel deposit, which was previously interpreted as outwash (White and Totten, 1979). The far southeast corner is occupied by till in the form of small end moraines and ground moraine. There is also a major alluvial valley deposit in this part of the study area. For the stack maps, tills are not differentiated into end and ground moraines, as on a traditional glacial-geology map. Instead, tills are divided on the basis of broad textural class where the unit "TG" represents a silt-rich unit found lake ward and "TE" indicates a clay-rich till found inland.

The major subsurface feature in the study area was the north-south trending buried valley, and it was the major challenge to mapping in the study area. A generalized stratigraphic model was developed from a wide variety of information, little of which could provide definitive guidance or insight. The first task toward modeling the stratigraphy was to use previous studies (White and Totten, 1979), wells, and base maps to develop a general reconnaissance model of the subsurface. This initial survey gave a basic sense of the sediments that might be encountered in the subsurface and, in particular, indicated that lacustrine sediments were an important component of this buried valley.

The surface was dominated by tills of Wisconsinan age that ranged in thickness from nearly 0 to 100 feet in the study area (Figure 11). A rough estimate of the thickness of this unit was estimated from the wells using the first gravel or sand as the boundary between the Wisconsinan tills and underlying sediments. However, the marker was very thin, absent, or ambiguous in many of the wells. The map was useful for estimating thickness within ±50% for defining stack sequences, but was interpreted with caution.

The next issue was determining a generalized stratigraphy below the major till unit. The task was highly interpretive. Water well records only provided basic lithology (clay, silt, sand, etc.) and gave no information on the geomorphic environment. Hence, a lithology of "clay" could have referred to till or lacustrine deposits (perhaps ice contact as well). The buried valley likely contained till, lacustrine, sand, and gravel deposits. However, the only direct evidence for the existence of lacustrine deposits was from an ODOT bridge boring south of the study area (Figure 12). This well showed a 25-foot thick layer that contained 0% aggregates at a depth of 40 feet, which is likely a lacustrine deposit. Even with this high resolution and quality evidence, a low aggregate till could not be completely ruled out for this layer, however.

Further information to aid interpretation was provided by a plot of the proportion of sand and gravel with depth for all the wells of the study area (Figure 13). Sand and gravel did not commonly occur in the upper 50 feet of the surficial deposits (where Wisconsinan till predominated). Below 50 feet, the likelihood of encountering sand and gravel deposits increased up to a maximum of 40% at a depth of 105 feet. Then the proportion of sand and gravel dropped off again, to around 0.25 from 120 to 145 feet. Finally, the proportion increased again, but was interpreted with caution, as a limited number of wells penetrated to this depth. (The trends in the proportion of sand and gravel had implications for geostatistical simulation and are discussed later). This plot was used to guide the placement of sand and gravel layers within the stack sequences.

This variety of information was interpreted, combined and simplified to create the "stack" for each of the polygons. A base stack was developed for the center (thickest part) of the end moraine and buried valley, which

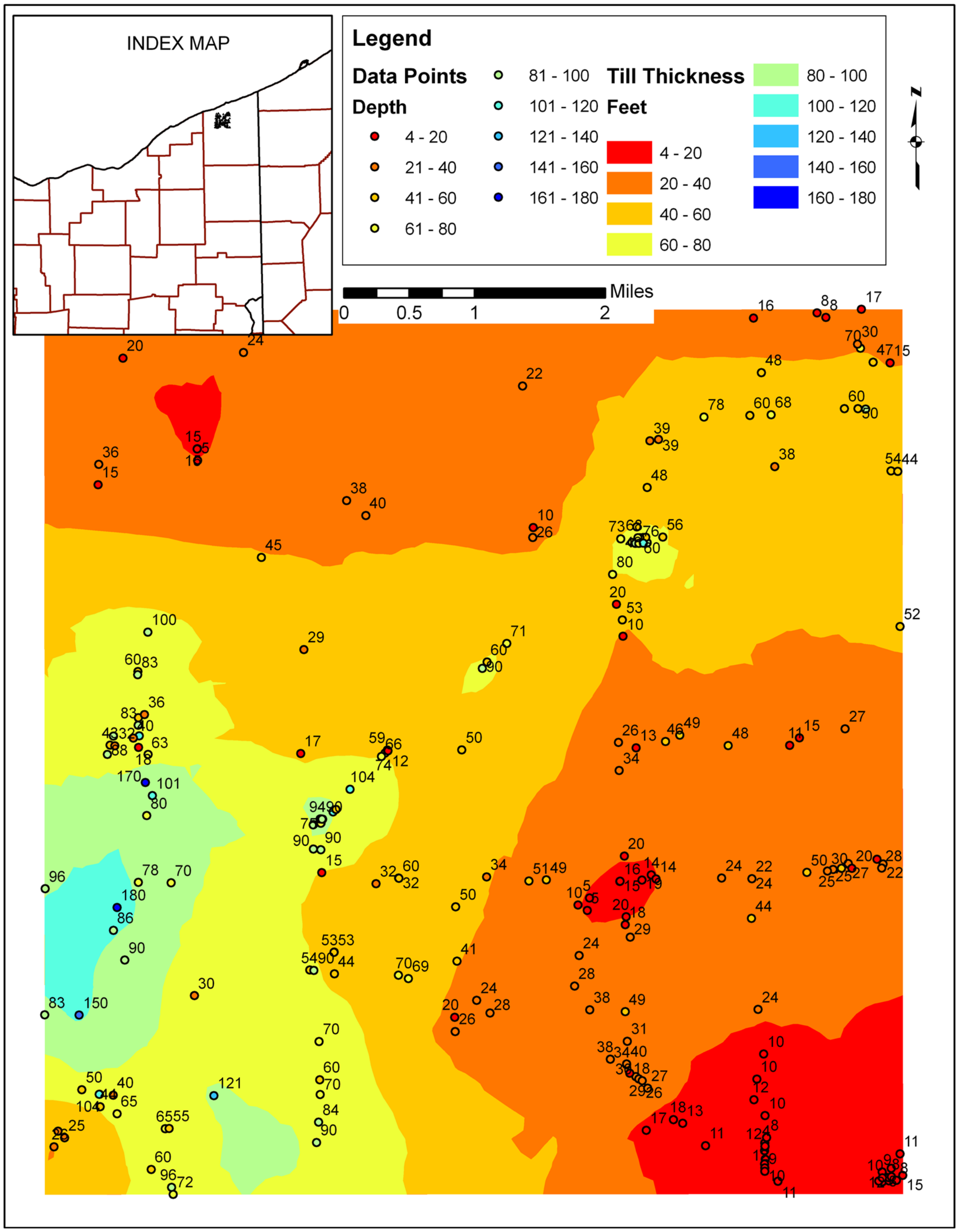


Figure 11. Estimated thickness of Wisconsinan till from kriging of water-well data.

was correlated outwards, eliminating lower units as the bedrock elevation increased. The data could have been interpreted and generalized in many ways, so several stack models were possible (Table 4). A range of models was given with varying degrees of complexity and interpretation. The layer of Wisconsinan tills (TG) was the most certain of the units and was used in all models. Model 1 was the most detailed and heavily interpreted version. The sand and gravel unit noted at the base of the tills was included in the second layer as (SG). Lacustrine deposits identified in the detailed bridge boring occurred below this unit. This was followed by a sequence of sand and gravel units estimated to be between 15 and 45 feet thick (SG3). This unit was based on information from Figure 13 (sections where the proportion of sand and gravel exceeds 30%). This was followed by another lacustrine unit, inter-

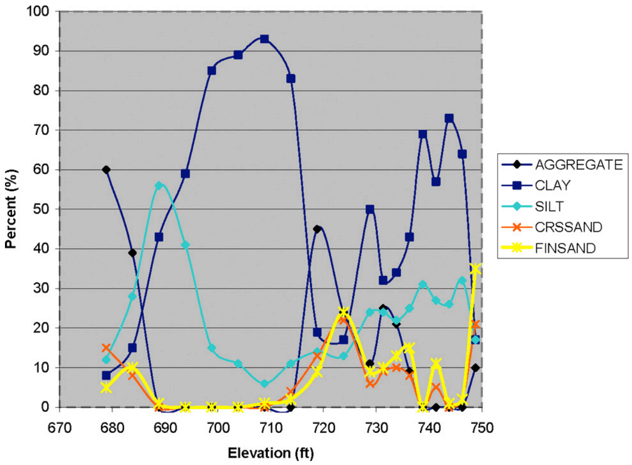


Figure 12. Texture analysis from an ODOT bridge boring close to the study area.

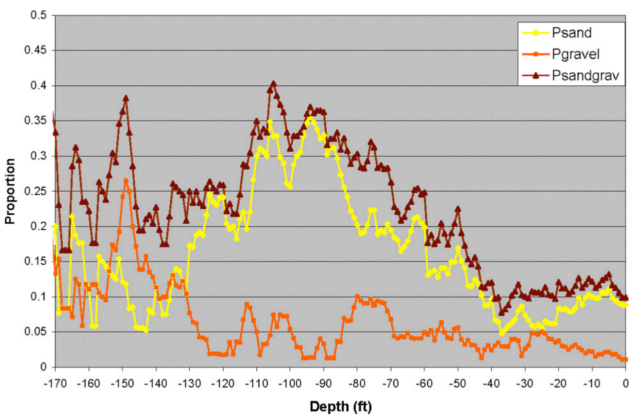


Figure 13. Proportion of sand, gravel and sand, and gravel with depth for all the wells in the study area.

Table 4. Potential stack models for the part of the study area with the thickest drift (an end moraine superimposed on a buried valley). The complex models (e.g., #1) might provide more information, while the simpler ones (e.g., #4) may provide a more reasonable picture of what is known about the geology in the area. The numbers represent thickness divided by 10, and are considered accurate to within 50%. Parentheses indicate that the presence of a layer is discontinuous between wells. Abbreviations are as follows, TG- Wisconsinan till with high silt content, SG- sand and gravel deposits, LC- silt and clay (generally lacustrine) deposits, CG- undifferentiated buried valley deposits with insufficient well control or extreme complexity that prevents differentiation of lithology.

	Model 1	Model 2	Model 3	Model 4
Layer 1	TG9	TG9	TG9	TG9
Layer 2	(SG)	(LC2)	SG4	CG16
Layer 3	LC2	SG3	CG11	
Layer 4	SG3	LC2		
Layer 5	LC2	CG9		
Layer 6	(SG)			
Layer 7	CG7			

preted from the drop in the proportion of sand and gravel. A buried till could was a possible alternate interpretation for layer 5. The water wells indicated a dominance of clay for this depth, but there is no information on aggregate content, etc., to show how this clay was deposited. Layer 6 in this model represented deep sand and gravels, which were mainly identified from detailed EPA site studies and descriptions in previous publications (White and Totten, 1979). This was a unit of pre-Wisconsinan till or sand and gravel that was oxidized and probably occurs near the bedrock interface. The rest of the sequence (on average, 70 feet of material) was essentially unknown because there were few borings that penetrated to this depth that contained material descriptions. This unit was designated as CG, which denoted buried-valley lithologies that range from clay to gravel. Model 2 was simplified by eliminating two units. The first sand and gravel (layer 2) of Model 1 was eliminated because it was often nonexistent or too thin to map. Also, the bottom-most sand and gravel of Model 1 was eliminated because evidence for its existence is questionable. Model 3 further simplified the model by removing the upper LC unit because it was confirmed at only one location. Likewise, the lower LC unit was eliminated, as it was purely an interpreted unit. There was little evidence to differentiate this interval between till or lacustrine deposits, so inclusion of this interval with the CG unit was justified. Model 4 represented the most conservative model. Here, everything below the Wisconsinan till unit was considered unknown. The justification for this approach was that the depth of occurrence and lithology of the buried-valley deposits below the till was essentially unknown and unmappable.

The choice of final stack model for the map was arbitrary. Decisions must be based on the judgment of the geologist, using a compromise between the limited available information (what can be justified on the data or evidence) and the need to communicate what likely would be encountered below the surface (based on geologic knowledge and interpretation). Model 4 was considered too simplistic. From Figure 13, it was clear that there was a nearly even chance of encountering a sand and gravel layer at depths ranging from 70 to 110 feet. The increased likelihood of sand and gravel layers over this depth range was not communicated in Model 4. Model 3 was also probably too simplistic. Here, the presence of sand and gravel was communicated, but no lacustrine deposits were designated. Model 2 contains all the major components that we expected to find in this buried valley. Two lacustrine deposits were designated, which bracketed the most probable stratigraphic position and thickness of sand and gravel. The stack model communicated the main idea of the deposit: tills underlain by buried-valley deposits that have more fine-grained materials than was typical for Ohio (due to the ice damming to the north). However, model 1 was too detailed and seems over-interpreted compared to the quality of the data. The bottom SG unit was only identified in a few wells and did not provide the user with particularly useful new information. The SG unit of Layer 2 was much more common in the well data, but thickness and depths are inconsistent.

In summary, it was difficult to correlate lithologies between wells with confidence. This was consistent with the results of variogram modeling, which indicated that spatial patterns were noisy at best (large nugget effect). Finding meaningful and reliable patterns between data points was a serious issue for both qualitative and quantitative mapping approaches.

Geostatistical Simulation

Example realizations, mode, and entropy are presented for each of the three variogram models to compare results. The results demonstrate the range of possible configurations (individual realizations) and the amount of spatial continuity using long and short autocorrelation ranges and the nugget effect. The results are presented as fence diagrams for an overview, and cross sections are provided for close inspection. An overview of the model domain, well data, and bedrock surface is found in Figure 14.

Individual realizations are presented in Figure 15, with two example realizations (of the 16 calculated) provided for each variogram model. Each obeys the data values, spatial structure, and histogram of the original data. Each realization is one possible configuration of the subsurface from a range of possibilities. There are clear differences between results. The 2000-foot range, large nugget effect model (Figure 15-A) produces realizations with a high amount of randomness. For example, simulated sand and gravel bodies contain many cells of clay, and regions of clay are “speckled” with sand and gravel

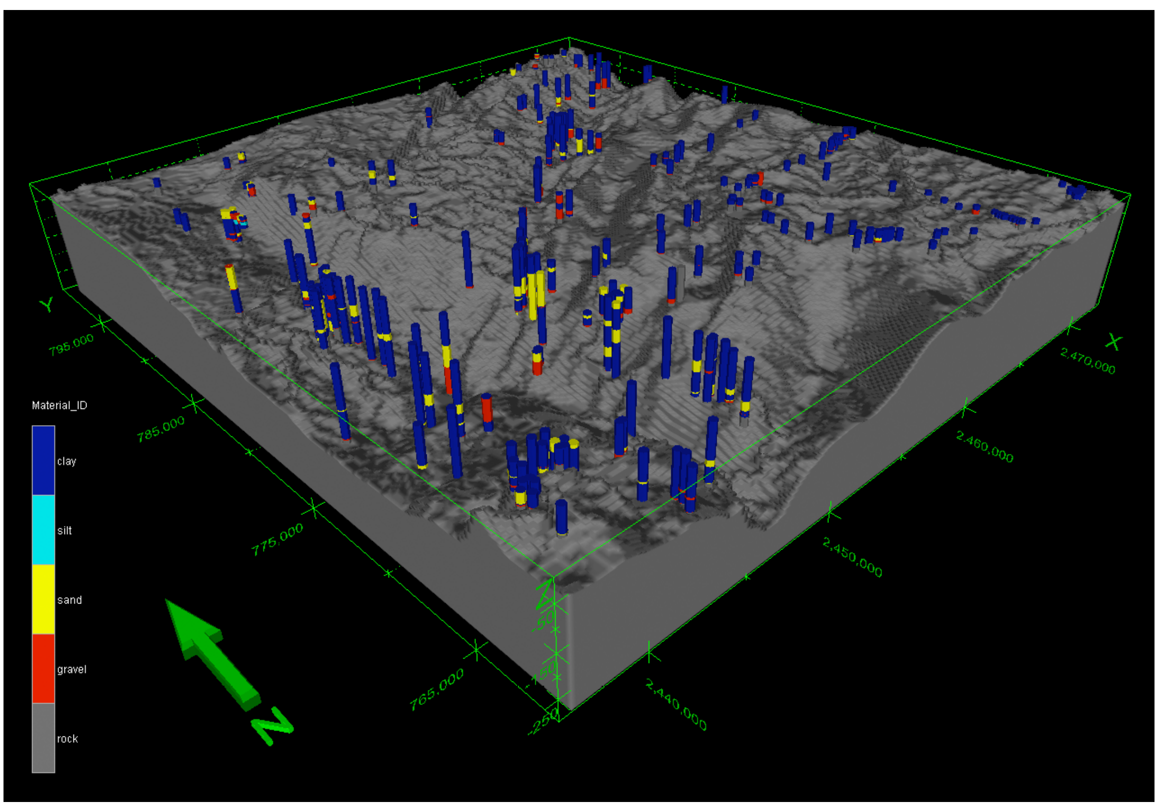


Figure 14. Overview three-dimensional model showing wells, their lithology, and the bedrock surface (in units of depth, not elevation).

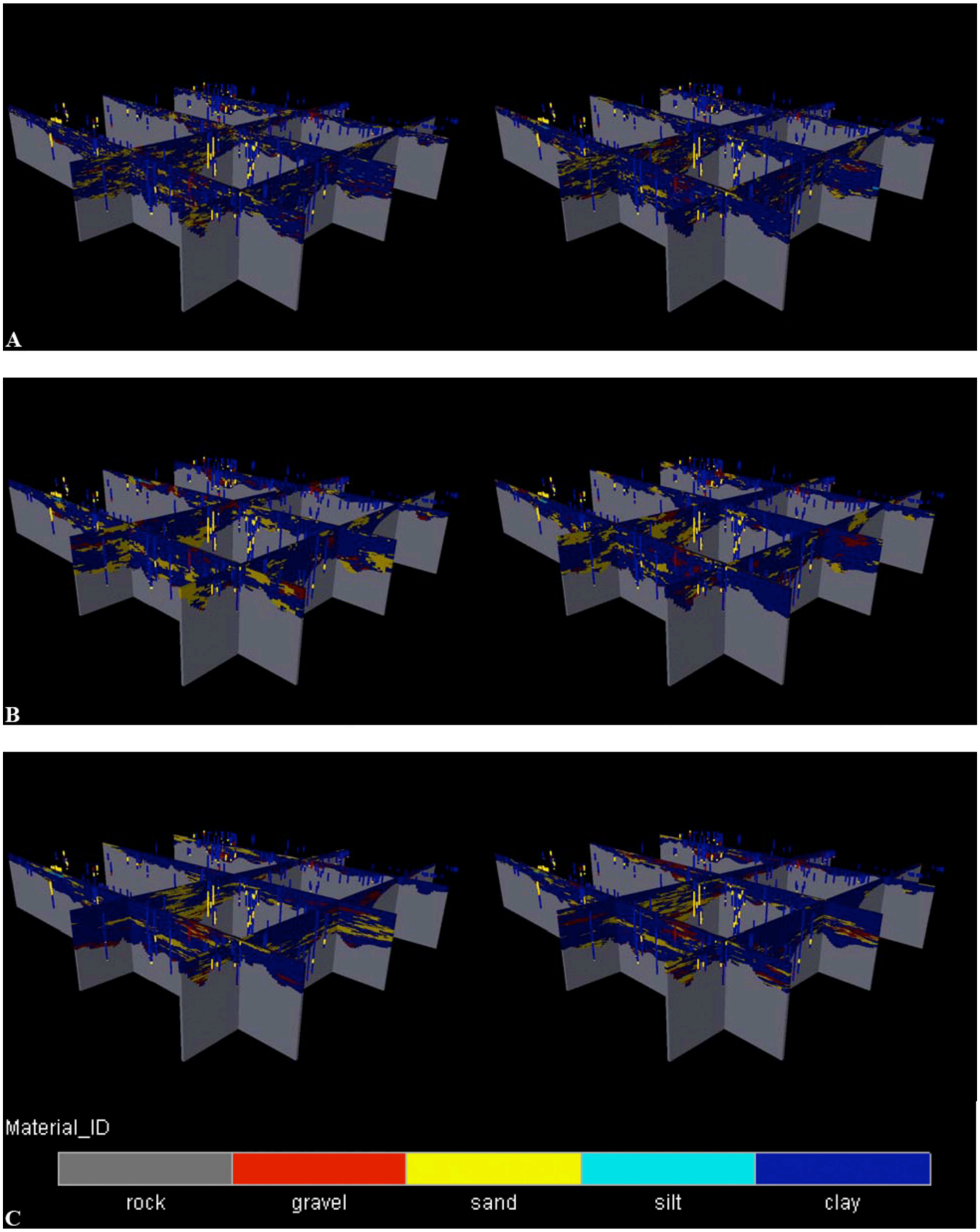


Figure 15. Example realizations for models. A, short-range variogram and high nugget effect; B, short range variogram with no nugget effect; C, long-range variogram with zero nugget effect. The spatial orientation is the same as for Figure 14 (north to the upper left corner).

cells. Such a model would have low flow continuity when modeling groundwater flow. Retaining the range (2000 ft) but removing the nugget effect produces a lithologic model with more continuous bodies (Figure 15-B). There is much more spatial continuity, and “speckling” is minimized. Extending the range to 5000 feet (Figure 15-C) produces elongate horizons of sand and gravel. Such results are visually pleasing, as they produce a layered look to the geology, which is compatible with stratigraphic concepts. However, this model is the least faithful of the three to the results of experimental variography. It represents the maximum extent of spatial continuity of sand and gravel bodies that could be reasonably interpreted from the data.

Inter-run variability between the short-range/high nugget and long-range models was investigated more closely by looking at an individual slice through the models. A cross-section (Figures 16, 17, 18 and 19) was chosen that has wells proximal to guide interpolation, but not within the displayed cells. This location was chosen in order to investigate the variation in simulation results where there was some well guidance, but also significant gaps. There, little commonality was found between single realizations of the short-range model and those of the long-range model (Figures 16-A and 18-A). As noted above, the short-range model had more interspersion between lithologies, and less contiguous sand and gravel bodies (increased influence of marginal probabilities over kriging weights). For the short range and nugget model, the data had limited influence on the results. For example, silt lithologies were found throughout the model, even though they were not present in any nearby wells.

The mode of the 16 runs (Figures 16-B and 18-B) also showed the increased randomness of the short-range model. The mode for the short-range model showed only one stable sand body, the result of nearby wells constraining the results. The lithology “clay” was the mode for most of the area. For the short-range model, the global proportions heavily influenced the results. The mode for the long-range model showed more sand and gravel in laterally extensive bodies. These sand and gravel bodies existed from simulation to simulation because surrounding wells contributed to the ccdf, so that each random-draw was constrained by data (the left terms of equation 2 had more influence than the global proportions).

The entropy results (Figures 17 and 19) further illustrated the differences between the models. For the short-range model, entropy results were mainly “granular” with little pattern, save for an area of low entropy in the upper right, where there was influence from a well. There were clear patterns in entropy results for the long-range model. The upper 60 feet or so generally had low entropy. The wells in the area were consistently clay, except toward the center, where there was less well control (allowing for more variation between realizations). The area of high entropy in the deep, central portion was produced

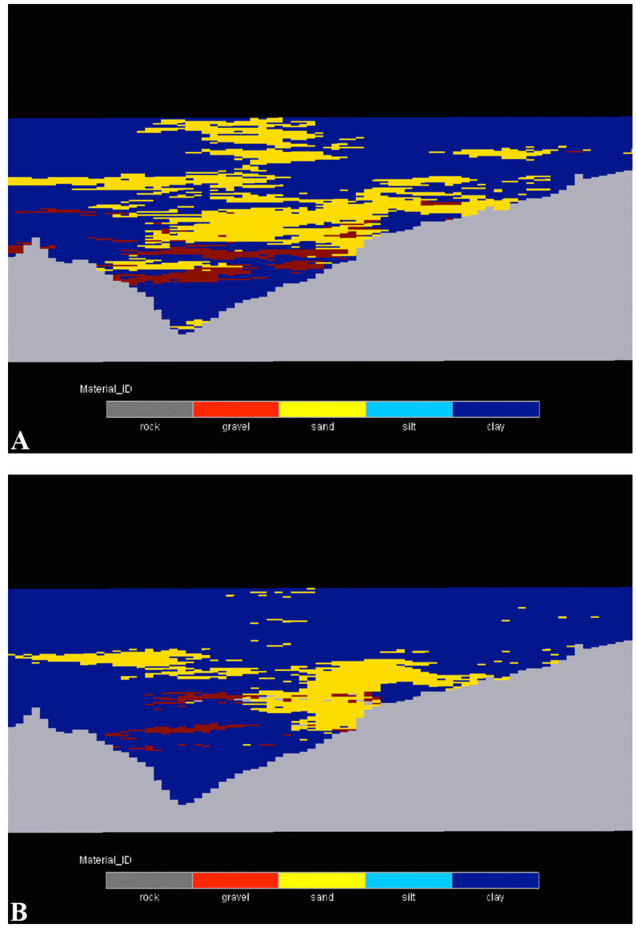


Figure 16. East and west-oriented cross section. Cross section is about 14,500 feet across and contains depths ranging from 0 (top) to 300 feet. Cross section results for long-range variogram with zero nugget effect. A, single realization; B, mode.

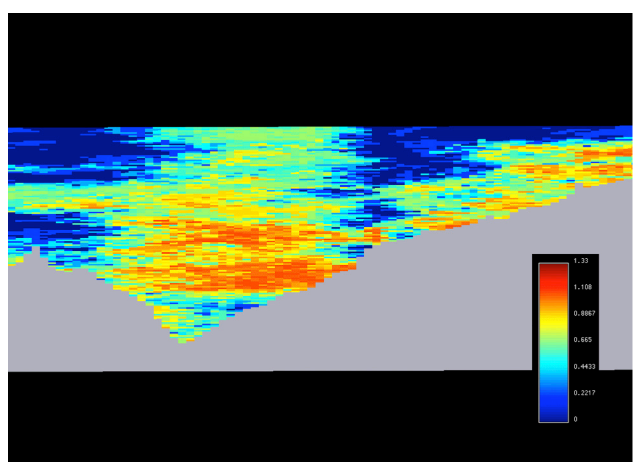


Figure 17. Cross section showing entropy results for long-range variogram. Orientation and dimensions are the same as Figure 16.

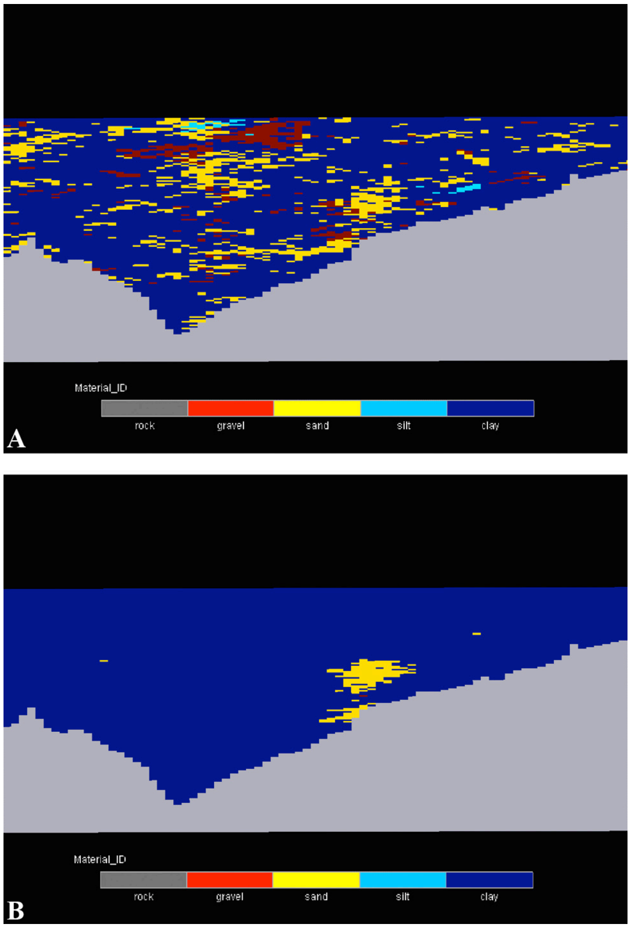


Figure 18. Cross section results for short-range variogram with nugget effect. A. single realization, B. mode. Orientation and dimensions are the same as Figure 16.

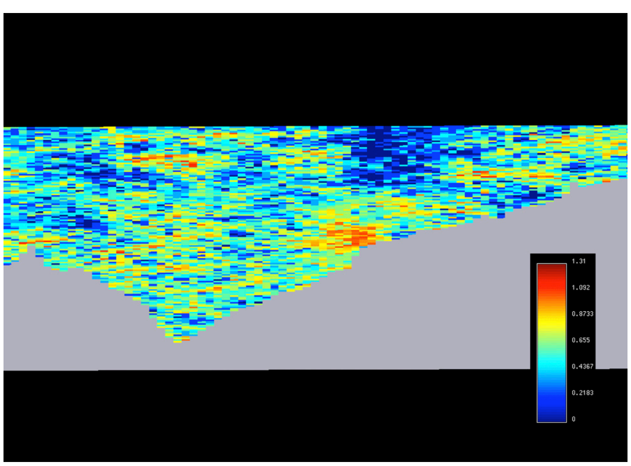


Figure 19. Cross section showing entropy results for short-range variogram with nugget effect. Orientation and dimensions are the same as Figure 16.

by alternation of sand and gravel between runs. True distinction between sand and gravel was questionable; an alternate modeling approach using hydrofacies (Ritzi et al., 2000) where the domain is divided into high and low conductivity units, which are modeled within stratigraphic units, is a better approach (Venteris, 2007).

CONCLUSIONS

The results raise important questions about the mapping of lithologies in buried valleys, particularly the feasibility of mapping and spatial modeling at 1:24,000 and more detailed scales. Most issues can be corrected with adequate well control coupled with geophysical studies, but this is not practical for the scales of interest due to the resources involved. Are the data good enough and is the geology predicable enough at this scale of interest to support county and regional-scale mapping of buried-valley deposits?

Of primary concern is the low horizontal continuity (lateral consistency) of lithologies between water wells for the current data set. This creates difficulties for both stack mapping and geostatistical simulation. Assigning a stack to a polygon implies that there is a predictable stratigraphy at that location. At simple locales (such as till over bedrock), the meaning of the stack is clear, and it is likely a reasonable prediction of the geology at that location. Tills can be correlated over large distances (Ehlers, 1996, chapter 9). Much of the surficial mapping work in Ohio is based on the correlation of tills. For this study (Figure 11), the scale of autocorrelation for predicting thickness of Wisconsinan till was on the order of 35,000 feet. The picture is less clear for mapping the lithologies of buried valleys. As seen in the water wells and simulation results, the depth of occurrence and thickness of sand and gravel bodies is highly variable. Some of this variability is due to the fact that lithologic layers encountered include both those that extend for long distances (stratigraphic units) and short distances (facies units). There are depth horizons where the occurrence of gravel is more probable, but these cannot be reliably traced from well to well. Such cases require careful consideration of what the stack sequence is predicting and communicating about the geology. In well-constrained situations, the stack can give accurate stratigraphic information, i.e., it represents a typical configuration of sediments for that polygon. When the driller puts in a new hole, he could expect a configuration and thickness of sediments similar to the stack sequence on the map. For more variable systems, an alternative, more probabilistic interpretation of the stack is warranted. In this case, the stack is an "average" sequence, which identifies the types of expected lithologies and their most likely vertical positions. For any location, however, portions of the stack may be absent and extra layers may be present. Wide ranges in thickness are possible as well. Comparison of the chosen stack model with the variability demonstrated in the geostatistical simulation results shows the difficulty in applying a stack sequence to complex buried-valley sediments.

Another important issue is the reason behind the poor horizontal correlation between wells. Several possible interpretations exist, the end members being:

1. The geology of this buried valley is predictable at this scale (well spacing), but the water well data are noisy and provide inconsistent information on stratigraphic and facies units. These complications mask the prediction of stratigraphy.
2. The water-well data are accurate, but geological variability occurs at scales well below the sample spacing (perhaps on 100m scales). The sample density is insufficient.

There is reason to believe that both case 1 and 2 are true. The water-well dataset is known to be noisy. Drilling crews with a wide range of geological training and experience produce these records as a legal requirement with the State of Ohio. Some water wells provide a very good approximation of local geology, while others contain serious errors in interpretation. Some of these errors were detected and fixed through the processing of the water wells into indicator variables, but certainly misidentifications and other blunders remain in the dataset. Past studies have suggested that case 2 may also be a contributing factor. Ritzi et al (2000) found that the range of autocorrelation for buried valley sediments is less than 1,000 feet. Therefore, very dense sampling is required (such as is conducted for site remediation studies) to make accurate three-dimensional models from experimental variography. An alternative is to use sources of more detailed information (geophysical profiles, outcrops) to develop models of spatial structure for geostatistical simulation.

REFERENCES

Bagley, C.T., 1953, Subsurface study of glacial deposits at Cleveland, Ohio: The Ohio Journal of Science, Vol. 53(2), p.65-71.

Brockman, C.S., 1998, Physiographic Regions of Ohio: Ohio Division of Geological Survey, Geologic Maps of Ohio series, accessed at http://www.dnr.state.oh.us/geosurvey/pdf/physio.pdf.

Brockman, C.S., Pavey, R.P., Schumacher, G.A., Shrake, D.L., Swinford, E.M., and Vorbau, K.E., 2004, Surficial geology of the Ohio portions of the Cincinnati and Falmouth 30 X 60-minute quadrangles: Ohio Division of Geological Survey Map SG-2 Cincinnati-Falmouth, scale 1:100,000.

Carle, S.F., and Fogg, G.E., 1996,Transition probability-based indicator geostatistics: Math. Geol., 28(4), p. 453-476.

Deutsch, C.V., 2000, Geostatistical Reservoir Modeling. Oxford University Press, 384 p.

Deutsch, C.V., and Journel, A.G., 1998, GSLIB: Geostatistical Software Library and User's Guide: Oxford University Press, 369 p.

Ehlers, J., 1996, Quaternary and Glacial Geology: J.S. Wiley, 578 p.

Journel, A.G., 1983, Nonparametric estimation of spatial distributions: Mathematical Geology, 15(3), p. 445–468.

Kempton, J.P., 1981, Three-dimensional geologic mapping for environmental studies in Illinois: Illinois State Geological Survey Environmental Geology Notes 100.

Ohio Division of Geological Survey, 2003, Shaded bedrock topography map of Ohio: Ohio Department of Natural Resources, Division of Geological Survey, BG3, 1:500,000 scale, with text.

Ohio Division of Water, 2005, Potentiometric Surface Mapping in Ohio: Ohio Division of Water Fact Sheet 05-65.

Powers, D.M., and Swinford, E.M., 2004, Shaded drift thickness map of Ohio: Ohio Department of Natural Resources, Division of Geological Survey, SG3, scale 1:500,000.

Powers, D.M., Laine, J.F., and Pavey, R.R., 2002 (revised 2003), Shaded elevation map of Ohio: Ohio Division of Geological Survey Map MG-1, scale 1:500,000.

Ritzi, R.W., 2000, Behavior of indicator variograms and transition probabilities in relation to the variance of lengths of hydrofacies: Water Resources Research 36(11), p. 3375-3381.

Ritzi, R.W., Dominic, D.F., Slesers, A.J., Greer, C.B., Reboulet, E.C., Telford, J.A., Masters, R.W., Klohe, C.A., Bogle, J.L., and Means, B. P., 2000, Comparing statistical models of physical heterogeneity in buried-valley aquifers: Water Resources Research 36(11), p. 3179-3192.

Shannon, C.E., 1948, A mathematical theory of communication: Bell System Technical Journal, v. 27, p. 379-423 and 623-656.

Smith, B., and Sandwell, D., 2003, Accuracy and resolution of shuttle-radar topography-mission data: Geophysical Research Letters 30(9), p. 1467.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database for *Survey Area, State* [Online WWW], accessed at http://soildatamart.nrcs.usda.gov.

Swinford, E.M., Pavey, R.R., and Larsen, G.E., this volume, New map of the surficial geology of the Lorain and Put-In-Bay 30 x 60 minute quadrangles, Ohio.

Totten, S.M., and White, G.W.,1987, Glacial geology of Mahoning County, Ohio: Ohio Division of Geological Survey Report of Investigations No. 133, 29 p., 21 figs., 3 tables, color map (1 inch equals about 1 mile), one sheet with text.

Venteris, E.R., 2007, Three-dimensional modeling of glacial sediments using public water-well data: An integration of interpretive and geostatistical approaches: Geosphere, v. 3.

Venteris, E.R., and Slater, B.K., 2005, A comparison between contour elevation data sources for DEM creation and soil carbon prediction, Coshocton, Ohio: Transactions in GIS, 9(2), p. 179-198.

White, G.W., 1971, Glacial geology of Trumbull County, Ohio: Ohio Division of Geological Survey Report of Investigations No. 80, color map (1 inch equals about 1 mile), one sheet with text.

White, G.W., and Totten, S.M., 1979, Glacial geology of Ashtabula County, Ohio: Ohio Division of Geological Survey Report of Investigations No. 112, 52 p., 22 figs., 2 tables, color map (1 inch equals about 1 mile).

Arkansas Geological Commission Template for 1:24,000 Scale Geologic Maps

By Jerry W. Clark and William D. Hanson

Arkansas Geological Commission
3815 West Roosevelt Road
Little Rock, AR 72204
Telephone: (501) 683-0152
Fax: (501) 663-7360
e-mail: jerry.clark@arkansas.gov, doug.hanson@arkansas.gov

This template (available at http://ngmdb.usgs.gov/Info/dmt/docs/hanson06a.pdf) was developed by the staff at the Arkansas Geological Commission for use in the layout of 1:24,000 scale geologic maps. The basic framework for this template was derived from various layout styles at DMT conferences, published U.S. Geological Survey geologic maps, and geologic maps produced by the Arkansas Geological Commission. Features included with this geologic map are the correlation chart, stratigraphic column, map unit description, cross section, joint frequency diagrams, symbols legend, state or agency seal, references, and disclaimer. Not all of these features occur on each of our maps, and layout size ranges from a minimum of 30 x 36 to a maximum of 42 x 36 inches for most maps.

The standardizing of fonts and creation of specialized symbol palettes is of great importance, as it will save much time and effort. Times New Roman font is used for text and symbols. Text is 12 point, while section headers are 16 point. The title block has various font sizes. The title itself is 36 point, the authorship and date published are 16 point, and the Director's name and agency are 18 point. The series and reference numbers are 10 point and located in the northeast corner of the layout. The disclaimer is 8 point and placed under the symbol legend if space permits. Revision dates are located in the lower right corner of the layout and are 10 point. The state or agency seal is placed in the extreme lower right corner of the layout and is 2 inches square. Formation contact lines are black and 1.0, and the formation symbols are 12 point. Colors used to signify different formations were adopted from the most current state geologic map, which was published by the U.S. Geological Survey.

Symbols used on the map are variable, but the decision was made to make them large enough to be easily found on the map. It must be understood that different symbol palettes will require different symbol sizes. This is another reason to create customized symbol palettes. Symbols for mines, quarries, and pits are 30 point. Strike and dip symbols are 60, while overturned strike and dip symbols are 30. The accompanying number is 8 point. Thrust fault, normal fault, syncline axis, anticline axis, and monocline axis lines are 2 point. Symbols accompanying synclines, anticlines, and monoclines are 30 point. Daggers for the thrust fault lines are size 10. The line that shows placement of the cross section is 1.5, and lettering is 16 point. The border of the layout has a line weight of 2.5.

In the future, we would like to incorporate digital pictures of interesting geologic features associated with the quadrangle. These images would be placed on the left side of the layout, space permitting.

This template was put together by J.W. Clark and W.D. Hanson for use in constructing 1:24,000 scale geologic maps at the Arkansas Geological Commission. By no means is this template engraved in stone. Much leeway is incorporated to give individual authors flexibility in designing their own layout

Geologic Map of the Ouachita Mountain Region and a Portion of the Arkansas Valley Region in Arkansas

By William D. Hanson and Jerry W. Clark

Arkansas Geological Commission
3815 West Roosevelt Road
Little Rock, AR 72204
Telephone: (501) 683-0115
Fax: (501) 663-7360
e-mail: doug.hanson@arkansas.gov, jerry.clark@arkansas.gov

One hundred and seventy-eight 7.5-minute (1:24,000) quadrangles in the Ouachita Mountain region and a portion of the Arkansas Valley region in Arkansas were mapped by Charles G. Stone (Arkansas Geological Commission) and Boyd R. Haley (U.S. Geological Survey) under the COGEO Map project (a precursor to the National Cooperative Geologic Mapping Program's STATEMAP component). The task of mapping this area took eight years and was completed in 1995. The maps completed during this project were compiled onto 1:100,000 USGS topographic quadrangle maps and then digitized and published (Haley and Stone, 2006). The map includes parts of the Fort Smith, Russellville, Conway, Searcy, Mena, Little Rock, DeQueen, Arkadelphia, and Malvern 1:100,000 quadrangles, and all of the Lake Ouachita quadrangle. Files from the ten digitized 1:100,000 quadrangles were merged in such a way as to produce seamless layers. Editing by the authors and additional staff members of the Arkansas Geological Commission was performed after each individual map was completed and after the compilation map product was produced. To make each map, heads-up digitizing was done over scanned base maps using ESRI ArcGIS 9.x software. Upon completion of the individual quadrangles, a new view was created and all files pulled together into a master file set.

Features included in these data sets are the surface geology (which includes formation contacts, strike and dip symbols, thrust, tear and normal faults, quarry, mine and pit symbols, igneous intrusions and igneous dikes), correlation chart, formation descriptions, and a few representative pictures of various geologic features encountered in the mapped area. The digital pictures are relatively new to our maps, and we hope they give the viewer a better understanding of rock types, structures, and outcrop appearances.

Formations that occur on this map from oldest to youngest are the Paleozoic Collier Shale (Oc); Crystal Mountain Sandstone (Ocm); Mazarn Shale (Om); Blakely Sandstone (Ob); Womble Shale (Ow); Big Fork Chert (Obf); Polk Creek Shale (Opc); Blaylock Sandstone (Sb); Missouri Mountain Shale (Sm); Arkansas Novaculite (MDa); Stanley Shale (Ms); Jackfork Sandstone (IPj); Johns Valley Shale (IPjv); Atoka lower, middle, and upper (IPal, IPam, and IPau); Hartshorne Sandstone (IPhs); McAlester (IPma); Savanna (IPsv); Mesozoic Cretaceous igneous (Ki) and undifferentiated (Ku); Cenozoic Tertiary undifferentiated (Tu); and Quaternary undifferentiated (Qu). In some instances, the Polk Creek Shale, Missouri Mountain Shale, and Blaylock Sandstone have been grouped together. When this occurs, the units are grouped under the SO symbol.

Economic resources that occur in this part of the state are coal, natural gas, cinnabar, antimony, novaculite, tripoli, crushed stone, dimension stone, shale, slatey shale, sand and gravel, clay, quartz, barite, manganese, copper, lead, zinc, vanadium, columbium, titanium, molybdenum, soapstone, and water. Landslides, most of which are induced by human activities, are the main geohazard in this area of the state. Ground subsidence can be a problem where historic underground coal mining has occurred. Controlled lakes built by the Corp of Engineers in this part of the state have greatly reduced the potential for major flooding, though minor flooding can occur when heavy rains occur in smaller tributaries.

Users of this map are state, regional, and local planners; local, state and federal government agencies; explorers of economic minerals; risk assessors; and those directly involved with earth sciences. Copies are available from the Arkansas Geological Commission, Little Rock, AR. A .pdf image can be accessed from the AGC website (www.state.ar.us/agc/agc.htm), which is about 350 Mb.

REFERENCE

Haley, B.R., and Stone, C.G., 2006, Geologic map of the Ouachita Mountain region and a portion of the Arkansas Valley region in Arkansas: Arkansas Geological Commission Map DGM-OMR-001, scale 1:125,000.

ArcGIS Geodatabase Schema for Geologic Map Production

By Vic Dohar

Natural Resources Canada
601 Booth Street
Ottawa, ON K1A 0E8
Telephone: (613) 943-2693
Fax: (613) 952-0738
e-mail: vdohar@NRCan.gc.ca

PREFACE

A general understanding of the ESRI geodatabase model would be beneficial when reading this document. Please refer to the accompanying poster at (http://ngmdb.usgs.gov/Info/dmt/docs/dohar06.pdf) for discussion below and note that the geodatabase schema that is represented is a DRAFT version.

A GEODATABASE FOR MAP PRODUCTION

This geodatabase schema has been designed to facilitate and manage digital data for the publication of geologic maps. In a production scenario, a personal geodatabase is created by each cartographer on their workstation. Geologic and non-geologic digital data are stored separately in respective feature datasets. All feature datasets and the feature classes contained within are prefixed by a seven-character string that represents the publication series and number (e.g., OF04780 represents Open File #4780). This ensures that all the files, when transferred from a personal geodatabase to an SDE enterprise geodatabase, will be unique. All digital base data are preserved in the original Shapefiles and stored outside of the geodatabase.

CARTOELEMENTS FEATURE DATASET

The CartoElements feature dataset contains non-geologic features that are used in preparing a geologic map. These consist of features such as the map border, UTM grid, leader lines, cross-section lines, base annotation, and masking polygons used with ArcMap's advanced layer masking option. Eventually over time, some of these features like the map border will be replaced with ArcObjects or macros used in ArcMap.

GEOLOGY FEATURE DATASET

The Geology feature dataset contains all the geologic features on the map that are represented by points, lines, and areas, as well as the bedrock and surficial geological units. Also included are topology rules that define the relationship between these features, a polygon feature that represents the area of interest, and geologic annotation. At a future date, raster geologic datasets will be a part of this dataset.

The geodatabase provides many options for managing these datasets efficiently, such as incorporating a subtype field and topology and domain validation. The three features classes that represent geologic area, point, and line features all contain a geodatabase subtype field named CATEGORY. Each unique value or subtype code in this subtype field can be viewed as a separate individual feature class. This provides better data management because common attribute fields can be shared by all features, with the ability to assign a different domain to each of these attribute fields for subtype code in the subtype field CATEGORY. Therefore, the attribute field FEATURE is assigned to a different domain for each unique subtype code. For example, the faults and folds subtype has the domain Faults and Folds assigned to the FEATURE attribute field, respectively (see Figure 1). The Faults domain contains the valid fault types that can be assigned to this field for any given feature, and the same can be said for the Folds domain. This method of managing data exists in the GeologyPoints, GeologyLines, and GeologyAreas feature classes since most features, depending on map scale, can be represented in any of these states.

The GeologyUnits feature class contains the bedrock and surficial geological polygons. In this feature class, the field MAPUNIT contains the unique value for each polygon feature. Several related tables are used to define each of these unique values. The first is the UnitComposition table, where each unique MAPUNIT value is defined as a composition of one or more geological units represented in the fields VENEER (used mostly for surficial geology maps), UNIT_1, UNIT_2, and UNIT_3. The geological units in these four fields are then related to the Bedrock-Geology or the SurficialGeology tables, each of which

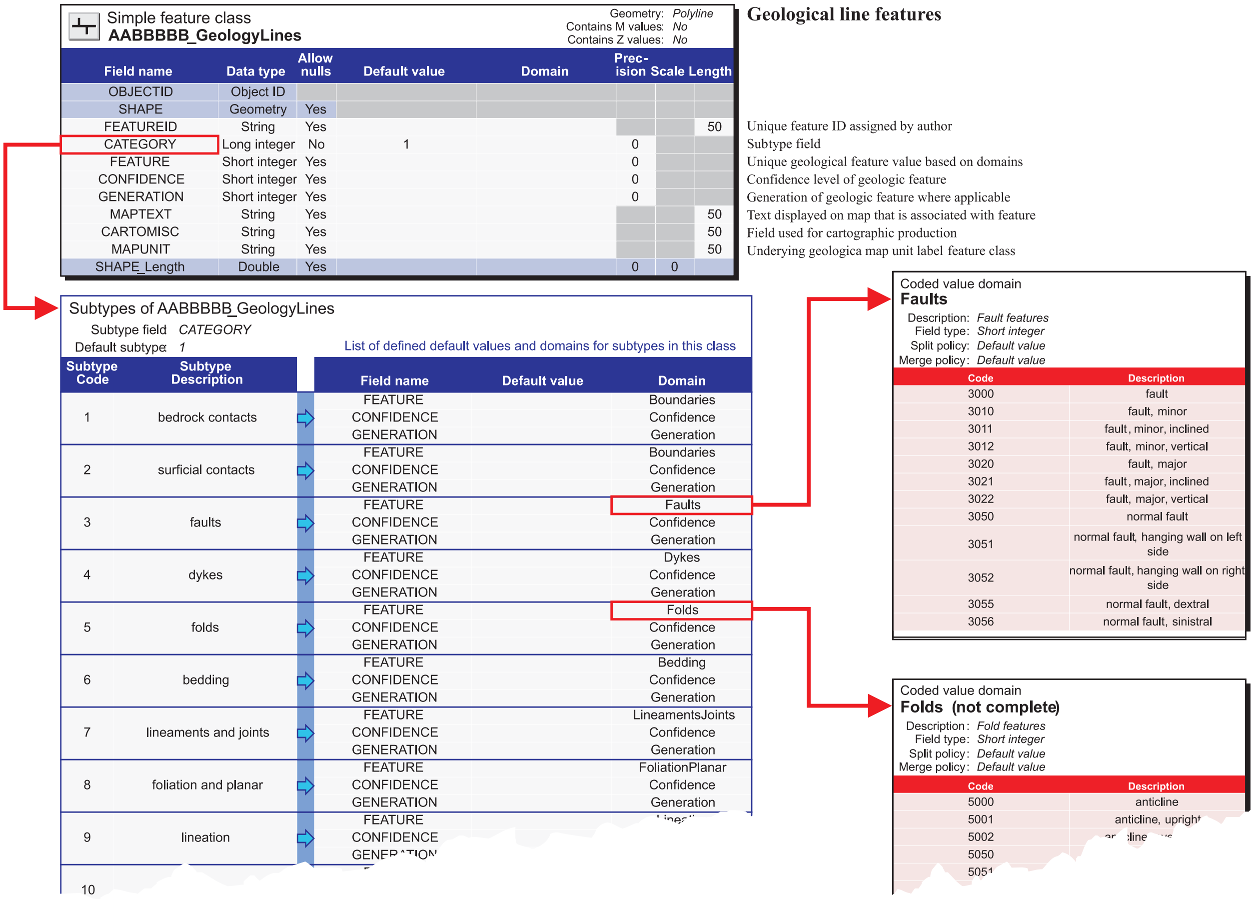


Figure 1. The CATEGORY field in the feature class is set as the subtype field, and its subtype codes are listed below. The subtype codes group geologic features in a common theme (e.g. faults, folds). For each subtype code, a separate domain listing all possible features of that subtype is assigned to the FEATURE field (e.g., the Faults domain is assigned to the FEATURE field for the faults subtype code).

further defines each geological unit and is comparable to the description that appears in the legend on the map. In addition, ID values in the field RELATION in the Unit-Composition table are related to the UnitRelation table that describes the relationship between one or more geological units (e.g., First unit is predominantly 60-80% coverage).

The MappedArea feature class simply stores one or more polygons that define the area of interest or geological study. The primary purpose of this feature is to ensure that all geological features exist within its boundaries, as part of the established topology rules. Also, the polygon features are used to define the FGDC element Data_Set_G-Polygon and for display in an index map on the final hardcopy version of the map.

The topology rules are basically those provided by the ESRI geodatabase model. As noted above, all geological features must exist within the boundaries of the MappedAreas feature class. In addition, it is also important to maintain coincidence of geological contacts, faults, and dykes to the outline of the geological units.

DOMAINS

Domains are used to constrain the values allowed in any particular field in a feature class or table. As mentioned above, each subtype value has a separate coded value domain that is assigned to the FEATURE field. Each of these domains contains the respective geologic features that have been compiled from the legend descriptions in GSC publications of the past ten years. Each geologic feature or description is assigned a code or a unique integer value in the 000's that corresponds to the subtype value (e.g., the subtype fault has a value of 3, and the Faults domain consists of coded values from 3000-3999). The use of domains aids in validating the feature attributes, ensures consistency and quality, and guarantees a homogenous dataset when combining the digital data from one or more geodatabases/publications.

CONCLUSION

The purpose of this geodatabase schema is to create a foundation for managing geological digital data in a map production environment. The ESRI geodatabase model provides the tools and validation methods to ensure a high quality output of data for geologic maps and the dissemination of digital data. These practices will continue to evolve as more datasets are incorporated.

REFERENCES

ArcCatalog and ArcMap, ESRI Inc., http://www.esri.com.

Microsoft Office Visio Standard 2003, Microsoft Corporation.

Zeiler, Michael, 2002, revised by Nichols, Greg, 2005, Geodatabase Diagrammer: an ArcScript available as a download from ESRI Support Center, http://support.esri.com.

Compression of Digital Orthophotography Collections: Factors to Consider in the Compression of Large Data Sets of Geospatial Imagery

By Deette M. Lund

Illinois State Geological Survey
615 E. Peabody Drive
Champaign, IL 61820
Telephone: (217) 265-5642
Fax: (217) 333-2830
e-mail: lund@isgs.uiuc.edu

INTRODUCTION

In early 2006 the Illinois State Geological Survey (ISGS) was scheduled to receive several large digital orthophotography collections. As part of the contract requirements, the individual images were to be compressed and made available to the public via the Internet. In December 2005 and January 2006, research was conducted regarding compression formats, compression software, and target compression ratios. During the course of our research, it became clear that the technology and standards involved with the compression of geospatial imagery were fast changing.

Image compression techniques have improved in the past few years. It seemed that whichever compression format we chose, it would yield visibly better results than those available five or even two years ago. The visible results of image compression are only part of what needs to be considered when making decisions regarding the compression of geospatial imagery. Metadata is also very important to geospatial imagery. Information such as pixel size, geographic location, and the coordinate reference system are just a few of the critical pieces of metadata embedded in a compressed geospatial image file that GIS applications need to properly display the image. A metadata standard is necessary for the variety of compressed file formats to interact with the GIS applications appropriately. Metadata standards[1] for compressed geospatial imagery, in some compression file formats, were not fully established as of January 2006.

[1]By June 2006, the metadata standards issue involving JPEG 2000 had been resolved. According to the Open Geospatial Consortium the "GML in JPEG 2000 Inter-operability Experiment (GMLJP2)" initiative has been completed. Currently it appears that all of the phases have been fully developed. Time constraints have prohibited any further research into these latest developments. Further research will be needed to determine what that means.

FORMATS

Two popular compression formats were included in the research: JPEG 2000 (non proprietary), and LizardTech's MrSID. A third compression format, Earth Resource Mapper's ECW format, was included in the initial stage of our research but was excluded due primarily to our long established relationship with LizardTech. The time constraints on our project did not allow time for building a new relationship with a different company.

JPEG 2000 Format

In 2004 & 2005 the JPEG 2000 compression format had become accepted as a standard by the International Organization for Standardization (ISO) and the International Electro-technical Commission (IEC). It became apparent during our study that JPEG 2000 was being developed in phases and that it was not fully developed (Morris, 2005). Several aspects of JPEG 2000 had been accepted as a standard by the ISO/IEC, but as of January, 2006 the geospatial aspects of the JPEG 2000 format were still in development and had not yet been approved as a standard. Another factor to consider with JPEG 2000 is that in 2003, according to Stuart Nixon, founder and CEO of Earth Resource Mapping (ER Mapper), there are at least three competing ways to store map projection information within a JPEG 2000 file, and our software developers use different methods (Thurston, 2003).

MrSID Format

LizardTech offers several compression algorithms within its latest upgrade of GeoExpress 6.0. Three that the ISGS considered were MrSID Generation 2 (MG2), MrSID Generation 3 (MG3), and JPEG 2000 (JP2). MG3 has improved compression capabilities. MG3 can compress in lossless format, 2:1 for black and white imagery and up to 6:1 in color imagery (ratios will vary from image to image). The lossy compression for MG3 is also improved, generating up to 50% better compression ratios (depending on the image) than MG2. Unfortunately not all GIS software packages have caught up with the MrSID technology, including software developed by Environmental Systems Research Institute (ESRI). Some of ESRI's GIS software packages are still not fully capable of using the MG3 format and the majority of the GIS user community in Illinois uses ESRI technology. For this comparison only MG2 and JP2 were tested.

ECW Format

ER Mapper's ECW format was not considered for this comparison due to a number of factors. The first was due to time constraints on our project which did not allow time to establish a new relationship with a different company. Secondly there were patent litigation[2] issues at the time of our research. Earth Resource Mapping (ERM), the parent company of ER Mapper, was in litigation with Galdos, the parent company of LizardTech, over issues involving patent infringements (Thurston, 2003). LizardTech started the litigation and claimed their patent had been infringed. The companies have been in litigation since October 1999 and although it appeared it would be resolved soon, the ISGS could not wait for an outcome. A third factor included several documents available on the Internet that report comparisons between ECW and MrSID formats (GIS Services, 2005; Warmath, 2004). Those comparisons did not promote ECW as the better format. In contrast to those comparisons, we did get some positive feedback about ER Mapper and the ECW format from the Digital Mapping Techniques 2005 forum about image compression.

SOFTWARE

Two popular software packages were included in our research: Leica's Erdas Imagine, and LizardTech's GeoExpress. A third software package, GeoJasPer, was initially included in the research but was excluded early on in the research due primarily to its lack of technical support services.

[2]As of January 31, 2006 the litigation between Earth Resource Mapping and Galdos was settled (http://www.ermapper.com/company/news_view.aspx?PRESS_RELEASE_ID=398). Earth Resource Mapping won its claims against Galdos, but too late to be considered by the ISGS.

Leica's Erdas Imagine

Erdas Imagine provides free image compression within its software application. One limitation that is that it only provides compression for files up to 50 MB for the MrSID format files. The 2005 USGS NAPP-DOQ files exceed 170 MB in size and the 2005 USGS Urban Area files exceed 70 MB in size. The size of the files ruled it out as an option before another, not so obvious, factor came into play, which concerned the fact that Erdas used LizardTech's Software Development Kit (SDK) in setting up its compression capabilities. The developer has already made some encoding decisions for the user. Erdas only allows the user to change some of the multiple encoding options that are available with GeoExpress.

LizardTech's GeoExpress

Prior to December 2005, the ISGS had used LizardTech's MrSID Geospatial Encoder to compress all existing ISGS orthophotography collections. We needed to factor in the cost of an upgrade if we were going to use LizardTech's software again. The ISGS hadn't kept pace with LizardTech's software upgrades. This was primarily due to fiscal constraints and low usage of the software by staff after the initial purchase to compress the 1998-2000 NAPP DOQ collection. The upgrades at the ISGS had stopped just short of LizardTech's decision to use "data cartridges" (a file that keeps track of the amount of imagery that has been compressed) as its new way to charge customers for compression. LizardTech's new GeoExpress 6.0 would be able to compress imagery using either MrSID or JPEG 2000 formats and offered an unlimited "data cartridge" at a set price.

GeoJasPer

Before the ISGS started the actual compression tests of the two formats it was determined that the project team would need to use a software application that had a technical support system. Then, if there were trouble with the software itself or how it was handling compressions, the ISGS staff could use the support service to troubleshoot and fix any problems. Through this decision it was decided that the ISGS would not use GeoJasPer since there was no technical support system.

COMPARISON CHART

A comparison chart between the two major compression formats was developed in an effort to organize the facts related to each factor in the decision-making process (Figure 1). This chart shows the factors an institution or agency should consider when making decisions about which compression format and software to choose. Some of the facts within the chart are time sensitive and may no longer be relevant.

End User

Another factor in our decision-making process that was not added to the chart was the end user. The ISGS had already "trained" its Clearinghouse user base to use MrSID compressed imagery. Using GeoExpress to com-

MrSID and JPEG 2000 Comparison*

FACTOR	MrSID (sid format)	JPEG 2000 (GeoJP2 format)
Software Choices	GeoExpress (LizardTech – PC, LINUX, SOLARIS* options) * note - have experienced trouble with Solaris installation. Also, instruction manual for command line encoding could include better examples. Erdas (LizardTech SDK - only useful on files under 50 MB)	GeoExpress (LizardTech) Erdas Imagine (software extension created from LizardTech Software Development Kit (SDK)) GeoJasPer (created from LizardTech SDK) ECW JPEG 2000 (ER Mapper)
Cost	GeoExpress 6.0 - Unlimited version = >$3000 or Data Cartridge Version = >$2000 per TB Erdas - different pricing available to each institution or agency	Erdas - different pricing available to each institution or agency GeoExpress - same as MrSID format costs GeoJasPer - free
Geography Markup Language (GML) standard	Follows the GML standard	Has GML in some cases. Still working on standardizations. Current status of future standardization is not clear. Currently there are at least 3 competing ways to store map projection information.
ISO Standard	No - because it is proprietary	Yes - but all phases not fully developed yet
ESRI Compatible	MrSID Generation 2 - Yes - but need to define projection or provide an .AUX file MrSID Generation 3 - Not in all cases	Yes - but potential issues with geospatial info - depends on code writers choice of where to store geospatial metadata (couldn't find any problems during limited testing)
Compatibility with other GIS Software Packages	MG2 - Majority of cases (with Plug-ins for a few) MG3 - Not in all cases http://www.gisservices.net/downloads/NYProgram.pdf (As of May 2004)	Not in all cases - and even then it might have problems with geospatial info
Web Browser (Free Viewer)	ExpressView Browser Plug-in	Yes - the ExpressView and a few others (some viewers are better than others)
Compatibility with Adobe CS2	Yes - by using MrSID Decode (free) - by using "Save as" in the ExpressView browser (be careful it only saves the image visible on the screen at the time but it will kick out a .TFW file if you save it to .TIFF format)	Yes - can place the image in Illustrator - can also use same "save as" method described in sid format No - can not open in Photoshop
Compatibility with Other Graphics Software	Yes - by using MrSID Decode (free) - by using "Save as" in the ExpressView browser (be careful it only saves the image visible on the screen at the time but it will kick out a .TFW file if you save it to .TIFF)	Many third party plug-in's available - some are free - some are free for the "lite" version and then you pay extra for more bells and whistles
Generates log file (for metadata and statistics puposes)	GeoExpress UNIX - Yes PC - Yes Erdas - Yes	Erdas - No GeoExpress - Yes GeoJasPer - Software not tested; No on-line information provided about log file generation
Target -vs- Actual Compression Ratio	GeoExpress - can be much different 12:1 can result in 9.64:1 Erdas - same as GeoExpress	Erdas - No log file to list actual compression ratio information GeoExpress - stays more on target (from existing tests) 12:1 is 11.94:1 GeoJasPer - Software not tested; No on-line information provided about generating actual compression ratio information
Batch processing	UNIX - Yes PC GUI - multiple file (not "true" batch processing) CMD - batch processing (similar to UNIX - not tested)	Erdas - possible according to help documents (not tested) GeoExpress GUI - multiple file (not "true" batch processing) CMD - batch processing GeoJasPer - possible according to on-line instructions (not tested)
Control over encode settings	UNIX - full (command driven) PC - full (can save established profiles), "pre-tuned" but user can alter all settings	Erdas - not as many options as GeoExpress GeoExpress - more control than Erdas, "not pre-tuned" like MrSID (can be good or bad thing), can't control gamma or weight GeoJasPer - only controls target compression - no other settings
Generates world file	UNIX - Yes PC - Yes	Erdas - No instructions available about generating a world file GeoExpress - Yes GeoJasPer - No instructions available about generating a world file
Customizable Metadata	UNIX - Yes PC - Yes	Erdas - No GeoExpress - Yes GeoJasPer - No

*All costs and statistics current as of January 2006

Figure 1. Factors considered in comparison of MrSID and JPEG 2000 Compression formats.

press the images into either MrSID or JPEG 2000 format would result in no changes to user instructions and or viewer downloads. Researching how well other software would provide a compressed image that would be able to use the established viewer and whether the compressed images from that software would load properly into ArcSDE was beyond the time frame of the project. These considerations would need to be re-evaluated under different funding sources and time constraints.

CONCLUSION

In a technically challenging process, we considered the pros and cons of each compression format and each software option. We chose GeoExpress and the MrSID Generation 2 compression format for compressing the large data sets that would begin arriving in spring 2006.

The ISGS chose 8:1 for the target compression ratio for the 2005 USGS Chicago Urban Area color orthophotography collection. Differences between the original and the compressed imagery at actual size are not detectable. Zooming in beyond the reasonable usefulness of the image, at pixel level, the user can see a few changes. Those changes appear to be slight shifts in color on a few of the pixel groupings, but they are not easily detected. The average size of the uncompressed file is 71.5 MB. To keep the download time to a minimum, an 8:1 target compression ratio produced files under 10 MB in size. The average actual compression ratio for the 4527 files in this data set was 8.3:1.

The ISGS chose 10:1 for the target compression ratio for the 2005 USGS NAPP-DOQQ grayscale orthophotography collection. There are little-to-no differences between the original and the compressed imagery at actual size. If users zoom in to 200%, "compression artifacts" (loss of edge detail and slight fuzziness) are visible. For the most part the "compression artifacts" in the compressed images do not affect the use of the images for research. The average size of uncompressed file is 177 MB (State Plane version). To keep the download time to a minimum, a 10:1 target compression ratio produced files around 20 MB in size. To date, we have compressed nearly one thousand of these State Plane version files. The average actual compression ratio for the ~900 State Plane version files that have been delivered is 9.6:1.

Due to rapid advances in standards and technology the facts are frequently changing in regards to image compression. Each institution or agency has its own particular factors to consider when dealing with image compression. The factors listed in Figure 1 should be used as a starting point or guide but the facts within the chart must be re-examined before deciding which formats and software to adopt.

REFERENCES

GIS Services, 2005, New York State Program - GeoExpress with MrSID: GIS Services, accessed at http://www.gisservices.net/downloads/NYProgram.pdf.

Morris, Steve, 2005, GML Content of JPEG 2000 format: North Carolina State University Library, December 30, 2005, personal communication (email).

Thurston, Jeff, 2003, Unlocking Data to Expanded Potentials: Interview with Stuart Nixon, Founder and CEO of Earth Resource Mapping, GEO Informatics Magazine, accessed at http://www.ermapper.com/document/doc.aspx?doc_id=79.

Warmath, Eric, 2004, State Mapping Advisory Committee Meeting Notes: Report on Image File Compression Software, April 15, 2004, accessed at http://www.nbmg.unr.edu/smac/apr2004.pdf

ADDITIONAL ONLINE RESOURCES

Brislawn, Christopher M., 2002, The FBI Fingerprint Image Compression Standard, accessed at http://www.c3.lanl.gov/~brislawn/FBI/FBI.html.

Digital Preservation Formats - http://www.digitalpreservation.gov/formats/index.shtml
Jpeg 2000 - http://www.digitalpreservation.gov/formats/fdd/fdd000143.shtml
Jpeg 2000 - http://www.digitalpreservation.gov/formats/fdd/fdd000140.shtml
MrSID Generation 2 - http://www.digitalpreservation.gov/formats/fdd/fdd000031.shtml
MrSID Generation 3 - http://www.digitalpreservation.gov/formats/fdd/fdd000184.shtml

Erdas Imagine Software Website
http://gi.leica-geosystems.com/default.aspx

ER Mapper Software Website
http://www.ermapper.com/

GeoJasPer Software Website (software function has changed since January 2006)
http://www.dimin.net/software/geojasper/

GIS Monitor - Newsletter
April 1, 2004
LizardTech and Galdos Take on JPEG 2000.
LizardTech/Earth Resources Mapping Lawsuit Judgment
January 20, 2005
LizardTech Introduces GeoExpress 5.0 with MrSID
December 15, 2005
Industry Survey: What was big news this year and what do you wish for next year?

Jakulin, Aleks, 2002-2004, Baseline JPEG and JPEG2000 Artifacts Illustrated, accessed at http://ai.fri.uni-lj.si/~aleks/jpeg/artifacts.htm.

Joint Photographic Experts Group - http://www.jpeg.org
Jpeg 2000 - http://www.jpeg.org/jpeg2000/index.html

Library of Congress, How to View - The American Memory Collections, accessed at http://memory.loc.gov/ammem/help/view.html

LizardTech - Press Room
March 23, 2004
LizardTech, Galdos Systems Collaborate to Develop ISO Standard for JPEG 2000.
May 3, 2004
LizardTech Unveils MrSID Software Developer Kit with JPEG 2000.

LizardTech – GeoExpress Software
http://www.lizardtech.com/

Morley, Karen, 2006, Avoid Pitfalls When Using JPEG 2000: GeoPlace.com, Tech Time Article, April 11, 2005, accessed at http://www.geoplace.com/uploads/FeatureArticle/0411tt.asp.

Pew Internet - American Life Project Demographics about internet use in America, accessed at http://www.pewinternet.org.

Wallace, Steve, 1999, Image Compression Software, accessed at http://www.directionsmag.com/features.php?feature_id=27.

Wikipedia
Jpeg 2000 - http://en.wikipedia.org/wiki/Jpeg_2000
MrSID - http://en.wikipedia.org/wiki/MrSID

Converting Adobe Illustrator Maps to ArcMap Format

By Jennifer Mauldin

Nevada Bureau of Mines and Geology
University of Nevada
Mail Stop 178
Reno, NV 89557
Telephone: (775) 682-8759
Fax: (775) 784-1709
e-mail: mauldin@unr.edu

BACKGROUND

The Nevada Bureau of Mines and Geology has developed a new cartographic production system for completion of geologic maps. Under this new system, all *new* maps will be completed using ESRI's (Environmental Systems Research Institute) ArcGIS Desktop software package—from initial digitization of lines through final layout design. Previously, a combination of software applications were used, which included Adobe Illustrator, ESRI ArcView 3.x, ESRI ArcInfo, Canvas, Microsoft Excel, and Avenza Map Publisher. Since the new cartographic production system is used for *new* maps, most geologic maps presently available at the Nevada Bureau of Mines and Geology are in Adobe Illustrator format and need to be converted to ArcMap format.

Various factors help us to determine whether a map will be converted to ArcMap format. We take into consideration time, effort, difficulty, and cost. For example, if an Illustrator map has been released as an Open-File Report (and therefore does not have a full office or field review), and a geologist decides to finalize the map for publication (full review) with minimal changes to the geology, we will opt to finalize in Illustrator rather than make the conversion. In instances, however, where funding for conversion is available or a request for conversion is submitted by geologists mapping in a particular area where a GIS (geographic information system) version will be beneficial, conversion of those maps under consideration will need to be implemented.

After a map is converted to ArcMap, final file types include an ArcMap document (.mxd), a geodatabase (.mdb), a topographic base map (.tif), and the digital file used for publication (.pdf). With the exception of the publication digital file, which is used for web and sales purposes at our agency, all those listed above are common GIS file formats for use with current versions of ArcMap. Other file types can be exported from these formats, which allows us to meet specific project requirements and provide digital data to those whose software applications are not compatible with the formats we typically provide.

There are substantial benefits to using the geodatabase format because all information in that format is bundled together as opposed to comprising a collection of shapefiles, coverages, dxfs, linked text documents, various projection files, etc., that make up the final map files. Although some of these files are generated as intermediate steps during the conversion process, they no longer need to be included in our final files after a map has been fully converted. The final digital files are few in number, organized, and pre-defined, which results in easy data transfer and viewing between colleagues, clients, and customers.

HOW THE ILLUSTRATOR MAPS COMPARE WITH THE ARCMAP MAPS

We have worked hard to make our ArcMap maps match the cartographic quality of our Illustrator maps. Improvements over the years to the ESRI software suite's cartographic functionality and presentation have made it possible to complete high-end cartographic products entirely in ArcGIS. Specific tools such as the Endpoint-Arc tool in ArcMap allow smooth digitizing of lines, while general improvements in the symbology options help to produce much more visually appealing products using ESRI software.

Specific cartographic differences we have observed between Illustrator and ArcMap maps include color display, font use, labeling methods, and difficulty in formatting the map layout in ArcMap. We have also dealt with differences between printer drivers after upgrading our large-format printer during transition to our new cartographic production system.

Color: RGB vs. CMYK

ArcMap and Illustrator maps differ in color display both onscreen and in print. ArcMap displays colors in RGB (red, green, blue) even though CMYK (cyan, magenta, yellow, black) color sliders may be used to enter the same percentages for each color value as used in the Illustrator CMYK color palette. The ArcMap CMYK color sliders simply allow a user to specify color values using the CMYK method, but do not actually *display* the map in CMYK color. When comparing a map printed from Illustrator with a version of that same map printed from ArcMap, it may appear that completely different colors were chosen to produce the two maps, even though Illustrator and ArcMap use the same CMYK color values. This is because the ArcMap map is actually displaying in RGB and, therefore, is printed in RGB even though the printer may be set to print using CMYK.

Fonts: Arial vs. Helvetica

Since the Helvetica font is not automatically installed on the majority of our geologists' computers, we decided to use Arial as the default font which does come automatically installed. This is due to the frequent file sharing with geologists and other cartographers at our agency who work directly from our map documents. We use Helvetica, however, on maps we decide to leave in Illustrator and especially those that cartographically began on a Macintosh. Often, Illustrator maps at our agency that have been started in previous years and are now nearing publication are Macintosh-based and were first created using the Helvetica font.

Labeling: Floating Text vs. Annotation

Labeling geologic units in Illustrator is simple; however, the labels are not georeferenced and are standalone text elements that are not linked to a GIS attribute table. To add a label, one needs only to type new text, or copy and paste an existing label and move it on top of its corresponding geologic unit. Labeling in ArcMap is not that easy. First of all, the labels are generated from the geologic unit attribute table, rather than being typed as "floating" text labels directly onto the map. After defining a label field in the symbology and turning on the label features option, the map is labeled but the labels are static and unselectable. To maintain control over label placements, we convert our labels to annotation after turning on the label features option. This creates a feature class that is added to the map document as a layer. The feature class has its own set of attributes in a table, and the features can be moved around on the map in editing mode.

Dealing with superscripts and subscripts within labels is also a challenge. Before converting labels to annotation, label classes are set up in the label properties, which allow VB Script expressions to be used for specifying superscripts and subscripts. Setting these expressions can be time consuming when many label classes contain superscripts and subscripts. Once label classes are set up and labels have been converted to annotation, proper placement of each label is needed. Due to the irregular-shaped geological units on a geologic map, annotation labels are not always placed in the best cartographic location and need to be manually moved to a better location. Additionally, some labels require leader lines and also need to be manually moved with a leader line assigned.

Layout: Graphical Interface vs. Technical Interface

To avoid using multiple applications for layout finalization, we now use ArcMap's Layout View rather than Illustrator. This not only allows the map to be completed in one document, but also enables an interactive view of the georeferenced data in the document's Data View by a click of a button. When Illustrator was used for final map layout, a separate application, for example ArcView 3.x, had to be opened in order to view the data interactively while displayed in its proper coordinate system.

The graphical interface of Illustrator contributed to the efficiency of layout finalization in our previous cartographic production system. However, the benefits of viewing spatial data in one map document and the efficiency and accuracy of making revisions to geology led us to finalize the map production in ArcMap. Even though ArcMap's Layout View is not as graphically oriented as Illustrator and often requires more steps to perform similar tasks, the overall map finalization process is much easier and organized using one application that provides all desired functionality.

Print Drivers: PS vs. RTL

As our agency moved toward our new conversion system, we purchased a 42-inch HP 5500 DesignJet PS3 large-format printer to be used as our draft plotter. After using the 42-inch plotter for our draft printouts and comparing various settings used, as well as previous drafts from old plotters, we discovered that using the RTL (raster transfer language) driver combined with the appropriate settings, instead of the Post Script driver, produced crisper lines, higher quality base images, and more accurate color.

CURRENT CHALLENGES DURING CONVERSION

Clean Up of Illustrator Layers

Illustrator documents are not always clean before converting to ArcMap. It is necessary to go through each layer and make sure map elements are on correct layers. We usually have to move misplaced elements back to their proper layers. Often, small elements such as unit labels, leader lines, strike and dip symbols and other small symbols are accidentally added to polygon, line, and other layers. Having map elements on incorrect layers can create problems when bringing the layers into GIS, such as causing polygon topology to fail to build and features to attribute incorrectly.

Preliminary Setup before Exporting

There are many steps involved in preliminary setup for conversion. Before an Illustrator file is ready to be imported into ArcMap, the user must add anchorpoints to the lines, simplify the anchorpoints on the lines using the straight lines option, and convert the file to a Drawing Exchange File (.dxf) before converting to coverage or shapefile. If these steps are not done properly, the coverage, shapefile or even DXF file will appear broken apart when viewed in ArcMap and be useless. Additionally, the user must make sure to join the Illustrator attributes (the layer names) with the coverage or shapefile so he or she can properly symbolize our ArcMap layers based on their attributes.

Line Clean Up, Building Polygons

Since we rebuild the polygons during conversion rather than converting the existing Illustrator polygons, we must clean the lines and check for potential problems that would cause the polygons to build incorrectly. Unfortunately, as an Illustrator map moves closer to finalization, it is much easier to make edits directly in the Illustrator document rather than going back to the original shapefiles or coverages that were used to import into the Illustrator document. This means that those original shapefiles and coverages become obsolete. Rebuilding polygons from our Illustrator line layer during conversion ensures that any edits made to the lines are reflected in the polygons, and our GIS data are accurate. However, this also means that we must perform clean up of overshoots, undershoots, and other errors that could cause the polygons to build incorrectly. To do this, we either use the ET GeoWizards tools (http://www.ian-ko.com/), Topology Rules in ArcMap, or the Advanced Editing tools in ArcMap to clean up our lines. Often a combination of methods is used during map conversion. This cleanup can require multiple iterations until all linework is properly closed and all polygons have properly built.

Re-labeling the Map

Although we found a quick way to convert Illustrator labels to ArcMap, it still requires clean up of duplicate labels or incorrect labels, which can be time consuming. The process involves exporting the Illustrator text labels to a point feature class that may be used to attribute polygons, and then further converting the point feature class to an annotation feature class, which requires manual cartographic placement of labels and leaders. The alternative to the quick method of generating labels is to manually select polygons on the map, attribute them, and generate labels, which are then converted to annotation. This method is ideal for maps that have simple geology; however, for very detailed maps, a judgment call should be made as to which method is more efficient.

Redigitizing Strike and Dip Symbols as Points

Since we use strike and dip symbols as point feature classes, which are symbolized and rotated within the map document, rather than graphical floating symbols as in Illustrator, the strike and dips are redigitized as a point feature class. They are then symbolized as a strike and dip cartographic marker symbol in the ArcMap document and rotated using the rotation tool with the geographic rotation option in the symbology window linked to the field in the attribute table that will store the rotation. As an alternative to redigitizing points, we are currently experimenting with exporting strike and dips to points to increase productivity of the strike and dip symbol conversion.

However, symbolizing the strike and dip point feature class in the ArcMap document leads to another challenge. When people request the files but are not capable of supporting our ArcGIS file types, they will only see the point feature class as a point, and *not* the symbolized strike and dip symbol, when they bring the point layer into their map documents. This is because the symbolization is stored in the ArcMap document that we supply. As a cartographic solution, we convert strike and dips to lines when using alternative file formats.

Although an entire new map can be completed from start to finish in ArcMap, the conversion process of an existing map from Illustrator to ArcMap is not as straightforward. As far as we are aware, ArcGIS simply does not have the capabilities to import a raw Illustrator file and generate a completely attributed map that is cartographically high in quality. In addition to an experienced cartographer, the conversion process requires multiple software applications and file formats to get the job done.

OUR SUCCESS WITH CONVERSION

Although we have only converted a handful of maps since developing our system for conversion, we have been successful in generating accurate and cartographically pleasing products that closely match the Illustrator versions. The balance between the extra steps involved in maintaining our high quality cartographic products when using ArcMap, and the ability to have georeferenced, attributed data all in one document, is reasonable. As the software continues to improve, our lives as cartographers will only become easier, which will allow us to produce more appealing maps, more accurately. The positives far outweigh the challenges that we have encountered, and we will continue to use ArcMap for map production.

APPENDIX A

Software used for conversion:
Adobe Illustrator 8, 10
Adobe Photoshop (minimal use) 7, CS
Canvas (for label conversion only)
ArcInfo Workstation 9.0, 9.1
ArcGIS Desktop 9.0, 9.1

Output devices currently used:
HP 5500 DesignJet PS3 (42-inch, dye ink)
HP 5500 DesignJet PS3 (60-inch, dye ink)
HP 5000 DesignJet PS3 (42-inch, dye ink)

Output devices previously used:
HP 2500 DesignJet
HP 755 Design Jet

Note: Avenza Map Publisher and ESRI ArcView 3.x not used during current conversion process.

Using GIS to Create and Analyze Potentiometric-Surface Maps

By Paul N. Spahr[1], A. Wayne Jones[1], Kelly A. Barrett[1],
Michael P. Angle[2], and James M. Raab[1]

[1]Ohio Department of Natural Resources
Division of Water
2045 Morse Rd., B-2
Columbus, OH 43229-6693
Telephone: (614) 265-6895
Fax: (614) 265-6767

[2]Ohio Department of Natural Resources
Division of Geological Survey
2045 Morse Rd., C-2
Columbus, OH 43229-6693
Telephone: (614) 265-6602
Fax: (614) 447-1918
e-mail: mike.angle@dnr.state.oh.us

INTRODUCTION

This paper describes the content of a poster display that is directly related to the oral presentation and paper of Jones and Barrett (this volume). Jones and Barrett discussed at length the GIS used to construct this poster as well as the database used to create the GIS layers. [GIS applications were based on Environmental Systems Research Institute (ESRI) software, including Arc Map, Arc Catalog, Arc Toolbox (Spatial Analyst, 3-D Analyst) and Arc Workstation (9.1).] Here, we will reintroduce their discussion and add some further explanation and details regarding the poster itself.

DISCUSSION OF THE POSTER

The poster (Figure 1) is available for download at http://ngmdb.usgs.gov/Info/dmt/docs/angle06.pdf, and contains a legend with a brief explanation and three panels. Each of the three panels depicts an identical base map of the study area overlain by various layers created by the process described in Jones and Barrett (this volume). The study area is a section of Darby Creek in western Franklin County, Ohio. A National and State Scenic River, Darby Creek provides an area of unique habitat to many endangered species of mussels and minnows. It is also an area that is undergoing rapid development pressure. Numerous parties, including planners, geologists, surface water ecologists, and fish biologists, were interested in determining the full impact of ground water flow on Darby Creek and a major tributary, Hellbranch Creek. There was particular interest in determining the gradient of the water table for the bedrock and unconsolidated (sand and gravel) aquifers, which are possible areas of ground water discharge to Darby Creek and Hellbranch Creek. Areas of recharge to the aquifers were of importance as well. It should be noted that part of the purpose was to create maps that would be easily understood by those with little geologic background.

The panel on the left shows the bedrock aquifer potentiometric-surface contours (arcs) superimposed over a topographic DRG basemap. This map helped portray the elevation, rough gradient, and direction of flow of the potentiometric-surface of the bedrock (limestone) aquifers.

The center panel shows the sand and gravel aquifer potentiometric- surface contours (arcs) superimposed over the same topographic DRG basemap. This map helped display the elevation, gradient, and direction of flow of the potentiometric-surface of the sand and gravel aquifers.

In most of the study area, the sand and gravel aquifer overlies the bedrock aquifer. There was an even distribution of water well log data points over the area that derived water either from the limestone aquifer or the sand and gravel aquifer. There was interest in the relative contribution of each aquifer to streamflow, and in the recharge areas of the aquifers.

The right panel shows a difference map that was used to compare potentiometric-surface maps. For each map, the shapefile was converted into a TIN model, which was then converted into a grid. Arc Map's Spatial Analyst was used to perform a subtraction between the two grids. A ras-

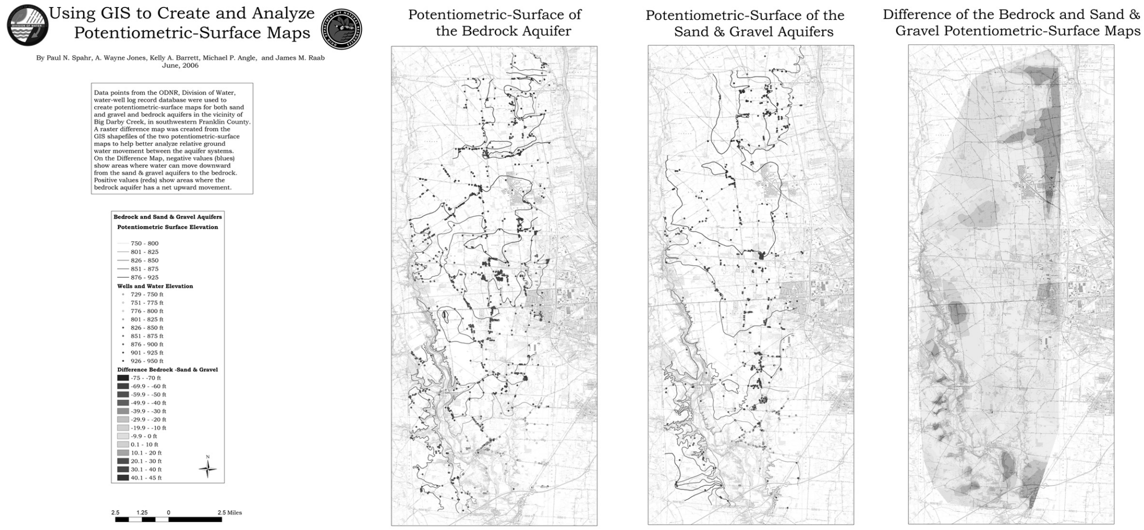


Figure 1. Poster "Using GIS to create and analyze potentiometric - surface maps".

ter difference map was then generated; this showed a value for the potentiometric head that indicates whether there was a relative movement of ground water from the sand and gravel aquifer into the bedrock aquifer or vice versa.

On the difference map, negative values (blues) show areas where water has a net downward movement from the sand and gravel aquifers to the bedrock aquifers. The blue areas are found on the uplands between Darby Creek and Hellbranch Creek. These areas tend to be recharge areas for both aquifers. Positive values (reds) show areas where the bedrock aquifer has a net upward effect, so water moves from the bedrock into sand and gravel units. These areas tend to be along the steep valley sides of Darby Creek and, commonly, are where the aquifers discharge. These small, steep tributaries of Darby Creek provide baseflow to the stream.

The information in this poster was widely accepted by the planners, biologists, and other stakeholders interested in the project. Readers are encouraged to view the poster in color at http://ngmdb.usgs.gov/Info/dmt/docs/angle06.pdf.

REFERENCE

Building a Water Well Database for GIS Analysis, by A. Wayne Jones and Kelly A. Barrett (Ohio Department of Natural Resources, Division of Water) DMT 2006, The Ohio State University, 125 South Oval Dr., Columbus, Ohio 43210. http://ngmdb.usgs.gov/Info/dmt/docs/jones06.ppt.

Updates to the Known and Probable Karst Map of Ohio

By Donovan M. Powers and Dennis Hull

Ohio Geological Survey
2045 Morse Rd. Bldg C-1
Columbus, OH 43229-6693
Telephone: (614) 265-6591
Fax: (614) 447-1918
e-mail: donovan.powers@dnr.state.oh.us

INTRODUCTION

The data shown in this map were created to establish a comprehensive statewide coverage of the known and probable karst features in Ohio (Hull, 2006). They show the subsurface locations of karsts and karst terrain. The project was conceived to help facilitate the location of a low-level radioactive-waste disposal site. This investigation was initially funded through a grant administered by the Ohio Low-Level Radioactive Waste Disposal facility Development Authority in 1997. The findings would be used to show the vulnerable karst areas and prompt the Development Authority to locate the waste disposal facility elsewhere. The primary source for the karst locations was the Ohio Cave Inventory provided by Dr. Horton H. Hobbs III of Wittenberg University. In addition, some investigation was conducted by utilizing geologic and soils maps as well as aerial photographs to find indications of karsts.

Thousands of sinkholes were inventoried but, for clarity, are not shown on the map. Field inspection of many indicated karst features allowed for documentation of more known karst locations and verified the mapping methods.

Probable karst areas were defined as those that: (1) lie within a half mile of a known or indicated karst location, and (2) are underlain by carbonate or gypsiferous bedrock with overburden generally of less than 20 feet of noncarbonate bedrock or unconsolidated material as shown by comparison of 7.5-minute bedrock-topography and bedrock-geology maps to surface topography. Known karst locations and probable karst areas on the work maps were digitized and are shown on this 1:500,000-scale map (Hull, 2006).

The map was updated by utilizing some of the datasets that the Ohio Division of Geological Survey has recently published as Geographic Information System (GIS) files. The criteria for carbonate or gypsiferous bedrock and for noncarbonate bedrock were queried out of the 2006 Bedrock Geology dataset, and like units were merged for simplification. The overburden of 20 feet or less criteria was derived from the Drift-Thickness 2004 dataset. This raster dataset was generalized and reclassified into greater than or less than 20-foot units. These two sets of data were then combined and added to the map.

PROCESS

The Known and Probable Karst in Ohio Map has gone through three iterations since its creation in 1999. The first version was created via utilization of Cadastral software (Figure 1). The next version was converted to a GIS format by way of Environmental Systems Research Institute's (ESRI) ArcInfo Coverage format (Figure 2). This conversion was also accompanied with updates and additions to the dataset. In this third version, we use newly created data layers and a more graphically oriented software package, the result of which is a more illustrative product (Figure 3).

The latest version incorporates data from three other products: Bedrock Geology, Drift Thickness, and the Glacial Margin extents from the Quaternary geology datasets. All of these layers were derived from sources within the Division of Geological Survey and are available as GIS datasets.

The incorporation of the bedrock geology units was a simple, straight-forward approach (Figure 4). The data are in a polygon vector format, so simply performing an attribute query on the bedrock units yielded those that meet the criteria conducive to karsting terrain.

The drift thickness data required a more extensive process (Figure 5). The data initially were in a raster format. For quality and illustrative purposes, the desired output needed to be vector. The process was to classify the raster into thickness of 20-feet or less and thickness over 20 feet. Next, the reclassified raster was converted to vector. This process caused stair stepping effect in the polygons and sporadic errors. A smoothing algorithm was used to better represent the areas. The smaller polygons were removed at a specific tolerance. The classified overburden layer was then merged with the carbonate or gypsiferous bedrock data. The dataset was closely checked against the original source layers and found to be topologically correct. The most time consuming portion of this maps creation was the clean-up and editing process.

A final addition used for assistance in visualizing the pattern to karsting was the glacial margin lines that were adapted from the 1999 Quaternary Geology dataset (Pavey and others, 1999) (Figure 6). Adding more basemap data helped make the final map more cartographically appealing. The cities, road networks, and hydrography were derived from the Ohio Department of Transportation datasets.

REFERENCES

Hull, D.N., 1999 (revised 2006), Known and probable karst in Ohio: Ohio Division of Geological Survey, Map EG-1 1999 (revised 2006).

Pavey, R.R., Hull, D.N., Brockman, C.S., and others with GIS compilation by Powers, D.M., 1999 (revised 2006), Known and probable karst in Ohio: Ohio Division of Geological Survey, Map EG-1 1999 (revised 2006).

Pavey, R.R., Goldthwait, R.P., Brockman, C.S., Hull, D.N., Swinford, E.M., and Van Horn, R.G., 1999, Quaternary geology of Ohio: Ohio Division of Geological Survey Map M-2, 1:500,000-scale map and 1:250,000-scale GIS files.

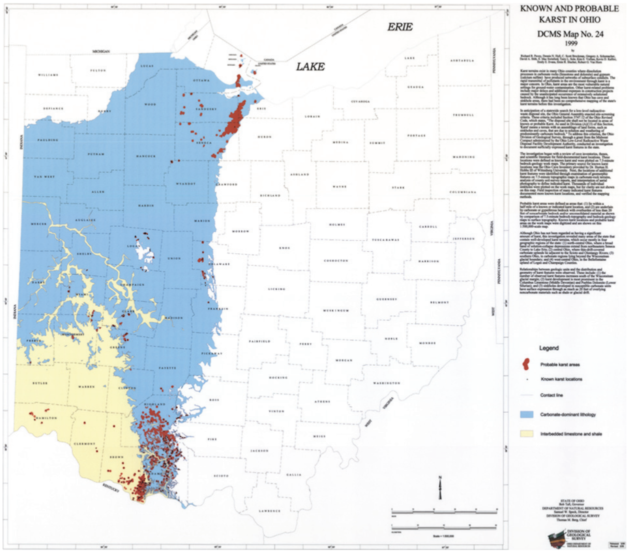


Figure 1. Version 1.0 (DCMS 24) of the Known and Probable Karst in Ohio, 1999.

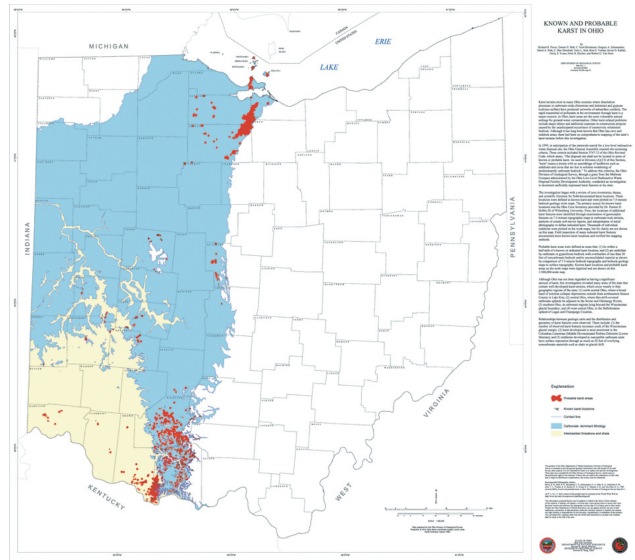


Figure 2. Version 2.0 of the Known and Probable Karst in Ohio, 1999.

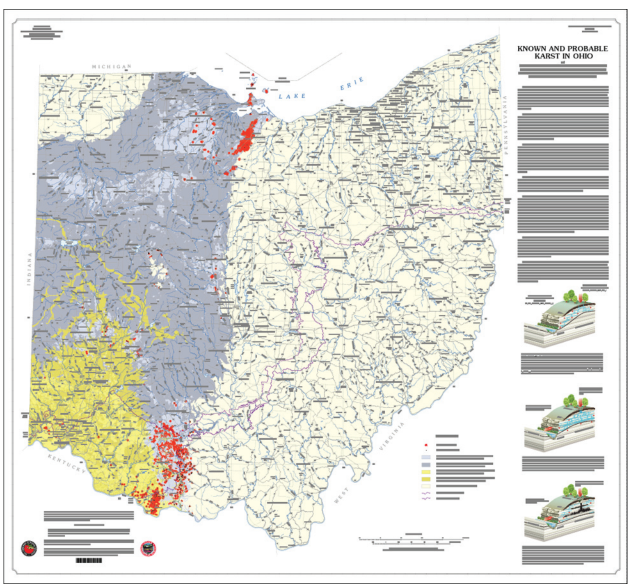


Figure 3. Version 3.0 of the Known and Probable Karst in Ohio, 1999.

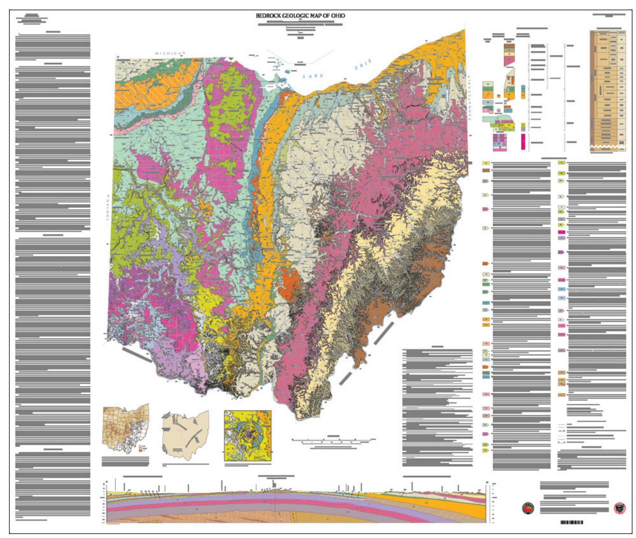


Figure 4. Bedrock Geologic Ma0p of Ohio, 2006.

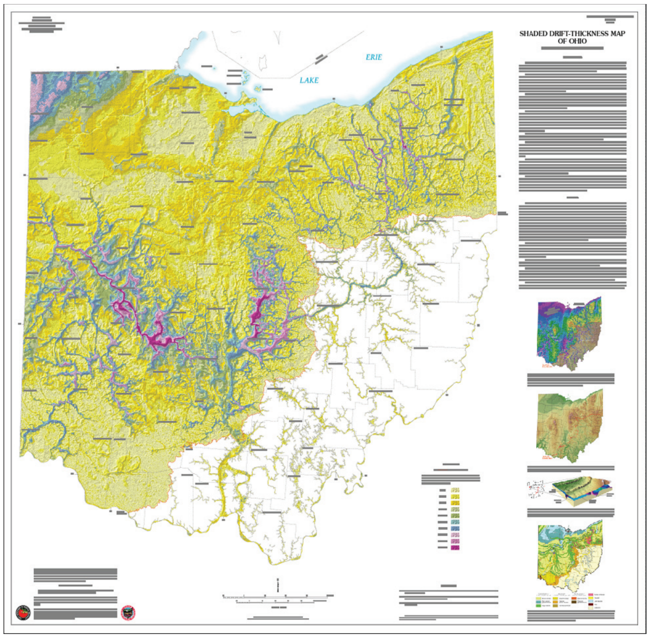


Figure 5. Shaded Drift Thickness Map of Ohio, 2004.

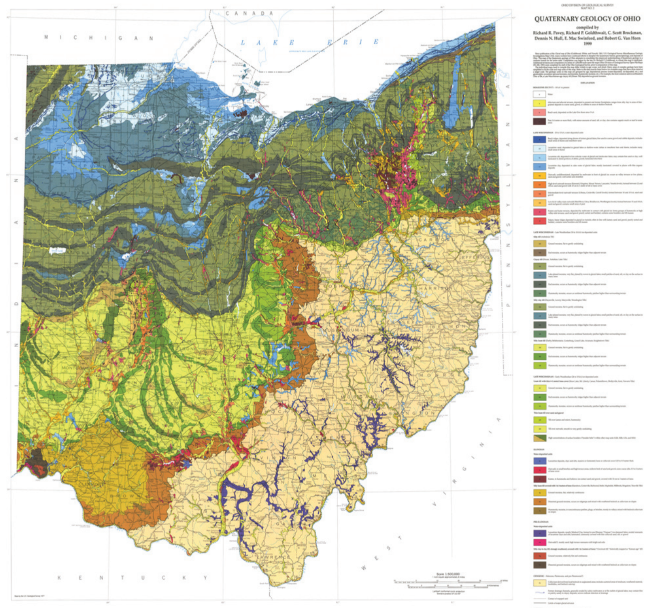


Figure 6. Quaternary Geology Map of Ohio, 1999.

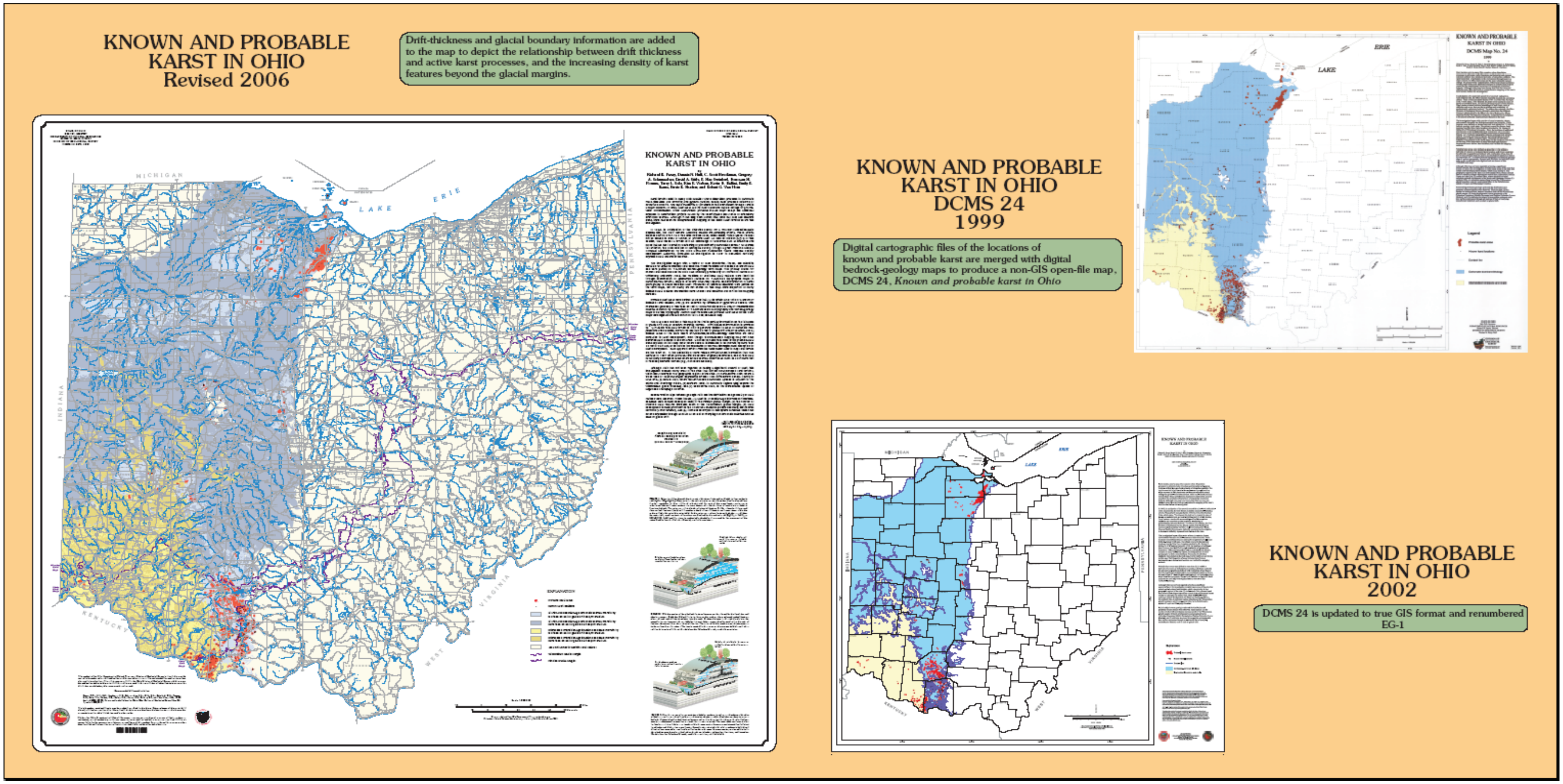


Figure 7. Poster of Known and Probable Karst in Ohio 2006 (as presented at the DMT meeting; see http://ngmdb.usgs.gov/Info/dmt/docs/powers06.pdf).

New Map of the Surficial Geology of the Lorain and Put-in-Bay 30 x 60 Minute Quadrangles, Ohio

By Edward M. Swinford, Richard R. Pavey, and Glenn E. Larsen

Ohio Department of Natural Resources
Division of Geological Survey
2045 Morse Road Bldg. C-1
Columbus, OH 43229
Telephone: (614) 265-6473
e-mail: mac.swinford@dnr.state.oh.us

ABSTRACT

A map depicting the surficial geology of the Lorain and Put-in-Bay 30 x 60 minute (1:100,000-scale) quadrangles has been produced by the Ohio Department of Natural Resources, Division of Geological Survey. Existing surficial maps at various scales document the uppermost surficial lithology of the area. The new map depicts underlying lithologies from the surface down to bedrock for use in geotechnical studies, land-use planning, and mineral exploration. To produce the new map, surficial deposits were mapped at 1:24,000 scale to create thirty-six 7.5-minute quadrangles, which were compiled digitally using GIS technology and converted into a full-color, print-on-demand, 1:100,000-scale, surficial-geology map. The map includes all or portions of Erie, Huron, Lorain, Lucas, Sandusky, and Seneca Counties in north-central Ohio. Data sources include field mapping, county soil surveys, Ohio Department of Transportation and Ohio Environmental Protection Agency boring logs, engineering logs, ODNR water-well logs, theses, and published and unpublished geologic and hydrogeologic reports. Map polygons were attributed using a stack-unit designator that indicates the thickness and stratigraphic sequence of major material units (i.e., till, gravel, sand, silt, and clay), from the surface down to and including the uppermost bedrock unit. Several regional material trends are apparent on the map, including large areas of lacustrine clay and silt landward of Lake Erie, the prominence of shallow bedrock that parallels the Lake Erie shoreline, a deltaic sequence deposited during higher levels of water of ancestral Lake Erie, the locally widespread and thick organic and marl deposits, and the expanse of Wisconsinan-age till that mantles the surface in most of the quadrangles. The text explains how to read the map, provides lithologic descriptions of mapped glacial and bedrock units, and offers other explanatory information. A GIS geodatabase contains spatial information on each polygon and data attributes of the stack units, all of which can be queried on the basis of material types and thicknesses for rapid generation of derivative maps. Potential queries for derivative maps might include isolating clay and silt deposits for the identification of potential geohazards, identifying sand and gravel deposits for aggregate exploration, or depicting areas of thick glacial till for the identification of potentially favorable solid-waste disposal sites. Mapping was partially funded by the U.S. Geological Survey, National Cooperative Geological Mapping Program, STATEMAP component. Digital compilation was made possible by funding from the Central Great Lakes Geologic Mapping Coalition.

INTRODUCTION

In 1997, the Ohio Department of Natural Resources (ODNR), Division of Geological Survey (OGS) began work on a long-range goal to produce reconnaissance-style three-dimensional surficial-geology maps for all of Ohio. The plan to attain this goal focuses on completing surficial mapping of major urban areas and highly populated corridors first in order that a majority of Ohio's 11.4 million citizens (2003 Census estimate) can benefit from modern surficial-geologic maps for land-use planning, resource exploration, hydrogeologic investigations, and geohazard identification. Less populated glaciated portions of Ohio will then be mapped followed by all of unglaciated Ohio. To date, this effort has resulted in the completion of three-dimensional surficial geology maps for 45% of the state's land area (354 of 788 7.5-minute quadrangles).

Existing Maps

Prior to recent mapping efforts, glacial-geology maps of Ohio were generalized, two-dimensional, geomorphically oriented products such as the 1:500,000-scale *Quaternary geology of Ohio* (Pavey and others, 1999) (Figure 1) and other published/open-file glacial-geology maps (generally at 1:62,500-scale) for 27 counties in north-

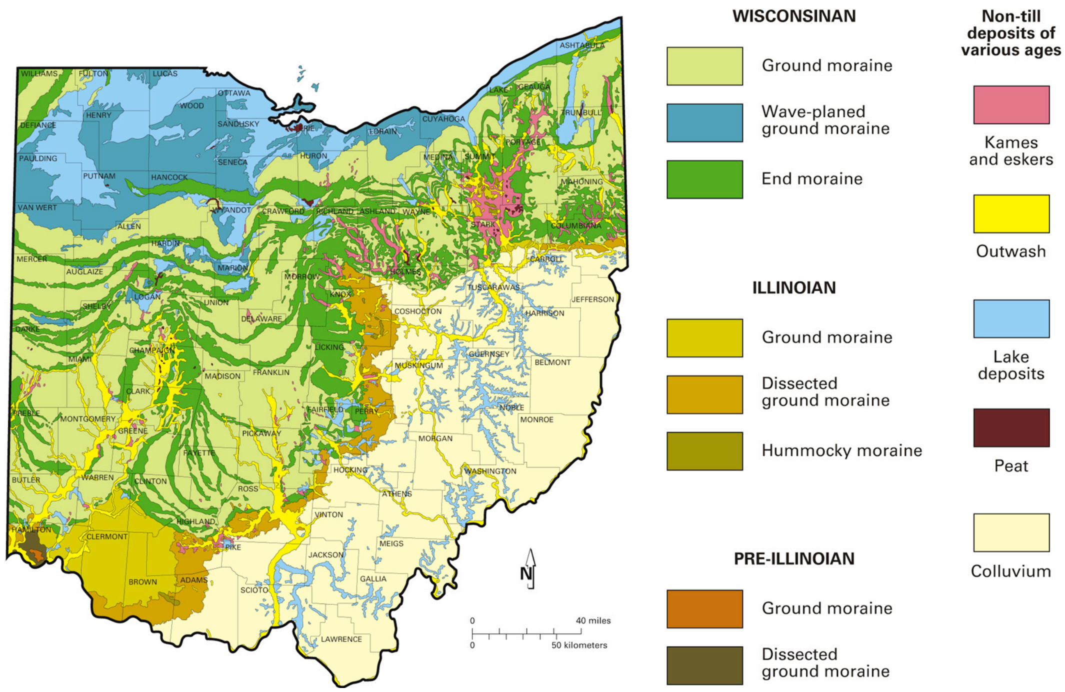


Figure 1. Map showing the glacial deposits of Ohio.

eastern, central, and southwestern Ohio (Figure 2). These older maps were constructed using a combination of field investigations, geomorphic analysis, and existing soils maps. Describing the entire surficial lithologic interval from surface down to top of bedrock was not attempted in this map set, as only the topmost unit was defined. However, existing glacial maps were used as a basis for selecting boundaries of the uppermost units during the remapping effort. While such maps can provide general information on the distribution of materials deposited by Pleistocene glacial and postglacial events, the maps are wholly inadequate for the characterization and assessment of unconsolidated materials at depth. The new surficial mapping effort adds the third-dimension component of variable lithologies at depth and their thicknesses.

New Glacial Mapping Program

In 1996, OGS conducted a survey of surficial-geology map users in Ohio to determine the kinds of map information they require for their needs. The majority of respondents to the survey questionnaire indicated a strong need for comprehensive, three-dimensional, surficial-geology maps that depict all deposits and their thicknesses down to and including the uppermost bedrock unit. In recognition of this need for more comprehensive surficial-geology data, OGS implemented a program to produce reconnaissance-style, three-dimensional, surficial-geology maps at 1:100,000 scale for the entire state. The mapping effort is based on a three-dimensional mapping method first implemented by the Illinois State Geological Survey (e.g., Berg and Kempton, 1988). The Illinois surficial mapping model used established glacial stratigraphic names (abbreviated) that are "stacked" as stratigraphic units would appear within the polygon they are defining. Ohio's surficial mapping effort modified the stack-unit concept to reflect lithologies of materials rather than glacial stratigraphic names and introduced additional constraints on unit thickness, allowing each area to be mapped down to the bedrock surface.

To date, reconnaissance-style, 1:100,000-scale surficial-geology maps have been completed for fifteen of the thirty-four 30 x 60 minute quadrangle areas of the state. An estimated 8.9 million Ohio citizens now have three-dimensional surficial mapping for various societal needs, including mineral-resource exploration, land-use planning, geohazard identification, and environmental protection. A database that contains the surficial-unit lithology, thickness, and distribution information on thousands of polygons shown on the map can be queried to produce derivative maps that identify geology of societal interest such as mineral resources or geohazards.

The new mapping program is largely funded by a tax on the mineral industries of Ohio, including oil and gas, with additional funding by the U.S. Geological Survey (USGS) STATEMAP program, the Central Great Lakes Geologic Mapping Coalition, and the U.S. Environmental Protection Agency Nonpoint-Source Pollution Program.

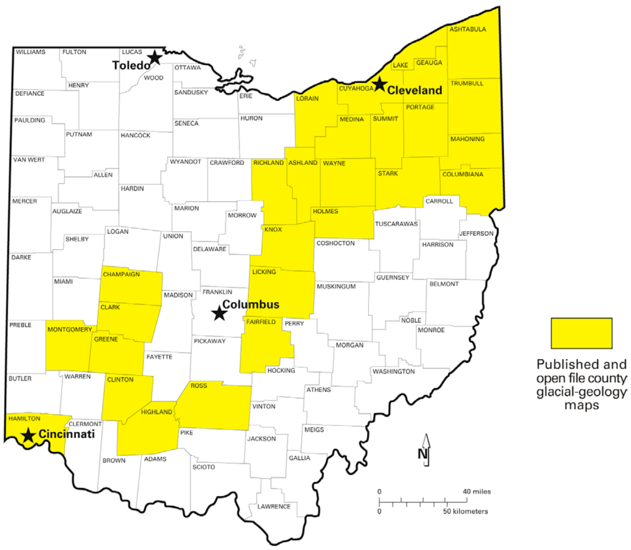


Figure 2. Map showing counties for which glacial-geology maps are available as published or open-file reports from the Ohio Department of Natural Resources, Division of Geological Survey or as part of Division of Water bulletins.

Purpose and Justification

Three-dimensional mapping of Ohio's glacial geology in urbanized or rapidly urbanizing areas (Figure 3) is a high priority for OGS. To date, OGS has completed 1:100,000-scale three-dimensional surficial-geology maps for the Akron, Canton, Cincinnati, Cleveland, Columbus, Dayton, Lancaster, Springfield, Toledo, and Youngstown metropolitan areas. These maps have been used by

1. private sand and gravel explorationists,
2. the Ohio EPA for waste-facility siting analysis and contaminated-site evaluations,
3. regional planning commissions for land-use planning,
4. colleges and universities as teaching tools,
5. private geotechnical firms for site evaluations, and
6. the Ohio Department of Transportation (ODOT) for shallow subsurface evaluations.

Major metropolitan areas and their surrounding interstate highway corridors (especially in glaciated areas of the state) are experiencing major economic development and related population growth. Land-use planning and industrial development in these corridors will benefit greatly from the three-dimensional mapping this overall effort will provide. OGS's long-range plan for surficial mapping in Ohio is to complete statewide mapping of the densely populated major metropolitan areas where most Ohioans live and work; surficial mapping of densely populated areas in glaciated Ohio is complete or in progress. When this phase of mapping is complete, the OGS mapping effort will focus on rapidly developing interstate highway corridors in glaciated portions of northern and western Ohio, such as the Interstate-71 corridor in northern Ohio, and the Interstate-75 corridor in western Ohio. After completion of the major portions of glaciated Ohio, mapping efforts will focus on the largely unglaciated terrain of southeastern Ohio, where thick deposits of outwash and landslide-prone glacio-lacustrine sediments occupy large portions of former and present-day river valleys.

METHODS

Construction of the Lorain/Put-in-Bay Map

OGS has compiled a map that shows the three-dimensional framework of the surficial geology, from the surface down to and including the uppermost bedrock unit, for the Lorain/Put-in-Bay 30 X 60 minute quadrangles (Pavey and others, 2005) located in north-central Ohio (Figure 4). Geologists at the OGS developed an easy-to-read format, described below, that depicts 1) the type of deposit, 2) the thickness range of the deposit, 3) the vertical sequence of deposits in the map area, and 4) the bedrock lying beneath the deposit.

Map Format Guidelines

- Map colors depict the uppermost continuous unit and are intended to assist users in visualizing the surface geology of the area (e.g., greens = till, reds and oranges = sand and/or gravel, blues and purples = silt and clay).
- Polygons or map-unit-areas define boundaries of the vertical sequence indicated by stack-unit descriptions that are composed of letters, numbers, and modifiers.
- Letters, numbers, and modifiers are arranged in stacks to depict the vertical sequence of lithologic units for a polygon. Simple abbreviations are used for ease of reading.
    - Letter abbreviations indicate lithology (e.g., SG = sand and gravel, T = till, L = silt).
    - Numbers indicate average thickness in tens of feet (e.g., 2 = 20 ft thick, + or - 50%).
    - Modifiers indicate aerial extent. A minus {-} sign following a number indicates the maximum thickness for that unit in areas like buried valleys or ridges. Parentheses () indicate that a unit has a patchy or discontinuous distribution in that map-unit area.

Data used to create the map were collected from numerous sources. The concentration of surficial data is greatest near the surface and decreases with depth. U.S.

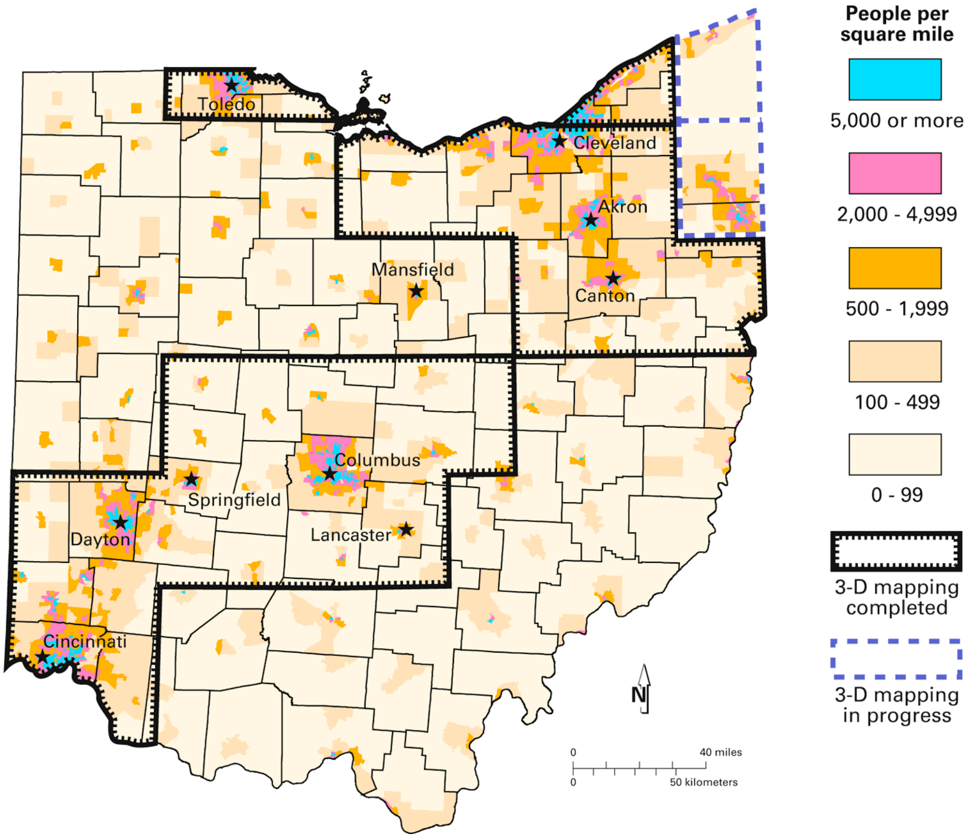


Figure 3. Map showing the population of Ohio by census tract and the outlines of areas with completed three-dimensional surficial mapping. Sources: the U.S. Bureau of the Census and the Ohio Department of Development.

Department of Agriculture, Soil Conservation Service maps, which describe the top 5 ft of surficial materials, provided an initial guide to map-unit area delineation. These areas were modified through interpretation of local geomorphic settings and other data that indicated a change in the type of deposit at depth, such as ODNR water-well logs, ODOT and Ohio EPA test-boring logs, engineering-boring logs, theses, and published and unpublished geologic reports, maps, field notes, and seismic-refraction profiles. These data also provided the basis for lithologic unit descriptions, which summarize, as accurately as possible, recognized associations of genetically related materials. The total thickness of surficial deposits was calculated by subtracting from land-surface elevation the bedrock elevation found on OGS open-file bedrock-topography maps, which are available for each 7.5-minute quadrangle in the map area. The bedrock units were summarized from OGS bedrock-geology maps, which are also available for each 7.5-minute quadrangle. Land-surface topography shown on the base map was prepared largely from data derived from the U.S. Geological Survey’s National Elevation Dataset.

The polygon and stack-unit information were hand-drawn at a scale of 1:24,000 on Mylar overlays registered to 1:24,000-scale 7.5-minute quadrangles. These Mylar maps were scanned, the line work was captured, and polygons were created. Stack-unit information that identifies the surficial geology from surface down to and including bedrock for each polygon was input into a geodatabase. Several iterations of quality control took place to ensure that line work between quadrangles and stack-unit assignments were edge matched. The color map consisting of polygons and stack-unit indicators was generated and included base map information and shaded elevation for the final map product. Other map elements include an explanation of how to read the map along with a schematic cross-section, detailed lithologic unit descriptions, references of sources used, a location map of the quadrangle, an index map that shows mapping responsibility, and a map color key.

RESULTS

Several regional surficial-material and bedrock-geology trends appear on the map. Large areas of lacustrine clay and silt deposited during higher levels of ancestral Lake Erie were mapped; they dominate the surface materials landward of Lake Erie in the central and western

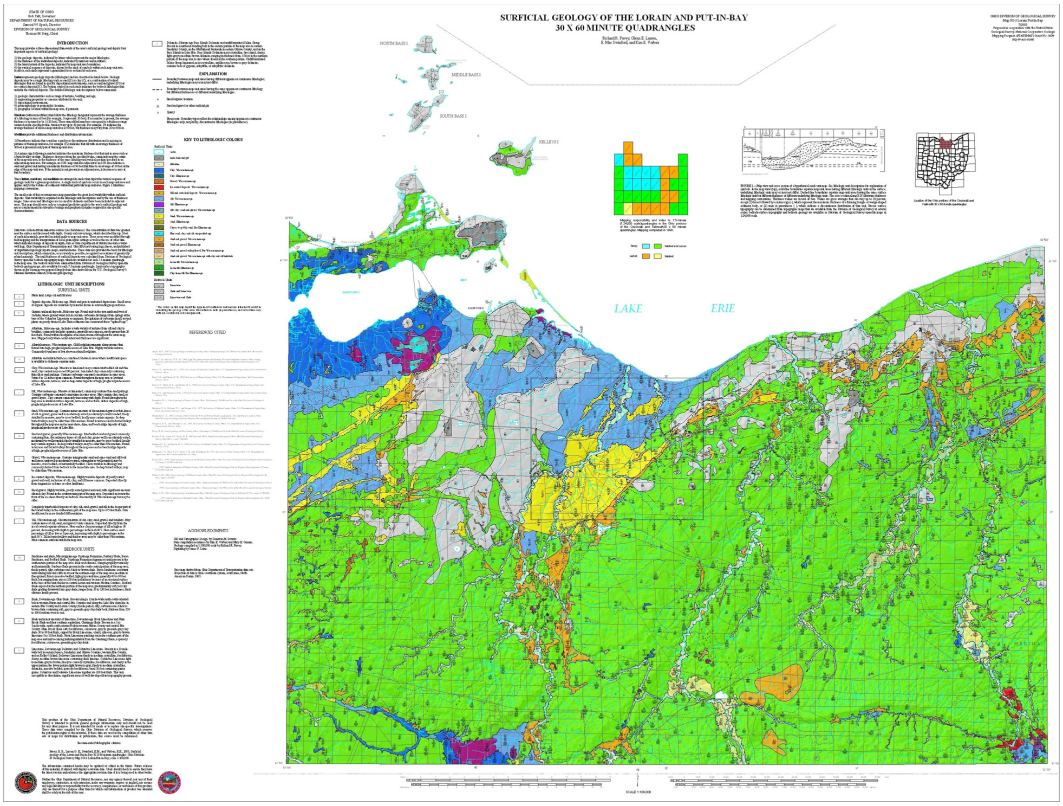


Figure 4. Surficial geology map of the Lorain and Put-in-Bay 30 x 60 minute quadrangles located in north central Ohio to be published as a full-color, print-on-demand paper map and released on CD-ROM disk and as a web-based interactive map.

portions of the map. Shallow bedrock parallels portions of the Lake Erie shoreline and ranges from economic deposits of limestone in the west to shale and economic deposits of sandstone in the east. A deltaic sequence of sand and silt deposited during higher levels of water of ancestral Lake Erie covers a large area in the central portion of the map. Locally widespread and thick organic and marl deposits, formed from the precipitation of calcium carbonate from local springs, were mapped in the northwestern portion of the map. Wisconsinan-age till, present as ridge and ground moraine deposits up to 120 ft thick, mantles most of the southern portion of the map.

Map Products

The final Lorain and Put-in-Bay 30 x 60 minute map will be released to the public in three formats: 1) a full-color paper format, print-on-demand, 1:100,000-scale, surficial-geology map; 2) a digital format on CD-ROM disk, which includes database files, base-map files, metadata files, and a PDF file of the original map; and 3) an Internet Map System product on the OGS website (http://ohiodnr.com).

Derivitive Map Products

The Lorain/Put-in-Bay 30 X 60 minute map is a digital product that can be manipulated to isolate various geologic components. Polygons and stack-unit information are in ArcGIS geodatabase file format and can be sorted by lithology and thickness to create derivative maps for a variety of uses. Figure 5 is a derivative product showing polygons that contain layers of sand, sand and gravel, or gravel with a thickness greater than 20 ft. Mineral companies could use this style of map to delineate areas that contain economic deposits of natural aggregate (sand and gravel) for potential exploration. Water-well drillers could use this map to delineate areas of thick, coarse sand and gravel deposits that may contain an abundant water supply.

Other derivative map products can be extracted from the digital data to suit many purposes. Examples include derivative maps that show areas of thick till, for potential placement of a solid-waste disposal facility, or areas of surface silt, clay, or organic materials that could indicate construction geohazards such as landslides and unstable near-surface materials.

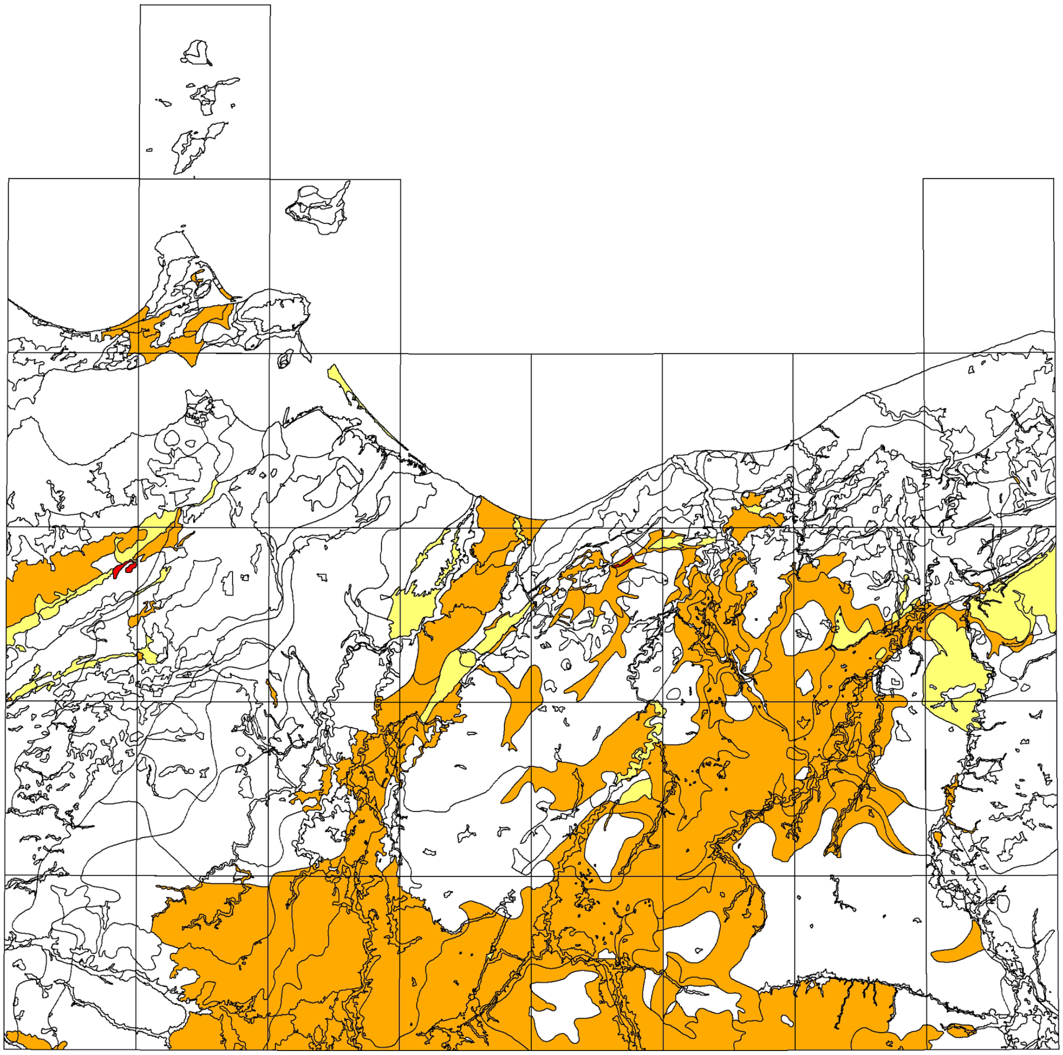


Figure 5. Digitally derived map (extracted from the original stack-unit information) showing areas that have sand, sand and gravel, or gravel with a thickness greater than 20 ft in the Lorain and Put-in-Bay 30 x 60 minute quadrangles.

REFERENCES

Berg, R.C., and Kempton, J.P., 1988, Stack-Unit Mapping of Geologic Materials in Illinois to a Depth of 15 Meters: Illinois State Geological Survey Circular 542, 23p., 4 pl.

Pavey, R.R., Goldthwait, R.P., Brockman, C.S., Hull, D.N., Swinford, E.M., and Van Horn R.G., 1999, Quaternary Geology of Ohio: Ohio Division of Geological Survey Map No.2.

Pavey, R.R., Larsen, G.E., Swinford, E.M., and Vorbau, K.E., 2005, Surficial Geology of the Lorain and Put-in-Bay 30 X 60-minute Quadrangles: Ohio Division of Geological Survey Map SG-2 Lorain/Put-in-Bay, scale 1:100,000.

Rapid Acquisition of Ground Penetrating Radar Enabled by LIDAR

By David Percy and Curt Peterson

Portland State University
Department of Geology
1721 SW Broadway
Portland, OR 97201
Telephone: (503) 725-3373
Fax: (503) 725-3325
email: percyd@pdx.edu

ABSTRACT

One of the most time-consuming aspects of collecting a Ground Penetrating Radar (GPR) survey is the acquisition of topographic data. To be useful, GPR data must be corrected for elevation, given that GPR "sees" only the data directly below it. We have developed a new method whereby LIDAR data can be used in conjunction with a Global Positioning System (GPS) for rapid acquisition of the elevation control that corrects GPR data for elevation changes. By collecting real-time GPS data and GPR at the same time, we can post-process these location data against a high resolution LIDAR data set and develop an elevation profile, which can then be used to offset the GPR data. This allows us to see features in the subsurface, such as groundwater surface trends or thickness of surficial deposits.

BACKGROUND

Ground Penetrating Radar (GPR) is a technique for surveying data in the shallow subsurface of earth materials (up to 40 meters deep, depending on frequency used, and characteristics of material) (Jol and Bristow, 2003). By generating an electromagnetic field and then recording reflections from subsurface materials with different transmissive properties, we are able to calculate the depth to subsurface objects of interest, such as the groundwater surface or the depth of surficial sediments (Peterson and others, in press). Each position along a line is occupied with a transmitter and a receiver at a fixed ground distance apart, depending on the frequency used. For example, a 100Mhz GPR survey will use an antenna spacing of 1m, while a 200 Mhz survey will use antenna separation of 0.5 m.

Every GPR survey that covers an area with varying topography requires a separate elevation survey that will correct for changes in elevation. As seen in Figure 1, this elevation correction can significantly improve the interpretation of GPR data. Here, we see how subsurface groundwater surfaces become horizontal, as expected, once the GPR data are corrected.

When collecting GPR data in the past, we surveyed in the topographic correction data using various methods that take nearly as long to collect as the actual GPR data. In 2004, our department purchased a cart that attached to a vehicle and allowed for transportation of the GPR, and data acquisition at speeds of up to 5 miles per hour. At this point, we needed a more rapid way to collect topographic data. Realizing that high resolution LIDAR data existed for our study area on the coast of Oregon, we decided to develop a methodology for topographic correction based on this.

LIDAR data are acquired via airborne surveys in which laser pulses are sent to the ground and the return times are collected, thus allowing for the collection of a very dense, high resolution elevation surface (Daniels, 2001). In some surveys, multiple return pulses are collected (fourth return) and can thus remove returns from higher elevation objects, such as trees, which results in a "bare earth" elevation surface. Since our collection typically occurs on roads, we can use "first return" data.

METHODS

A Global Positioning System (GPS) receiver is set in the dashboard of the collection vehicle. This allows for a reasonably good view of the satellites and, given that the GPS can receive through glass, the simultaneous collection of positional and GPR data. A cart is attached to the rear of the vehicle on which the GPR antennas are situated. Fiber optic cables connect the antennas to the interior of the vehicle, where the triggering and acquisition equipment are operated. An odometer wheel located behind the antennas triggers the computer, again via fiber optics, to transmit to the antennas controlling commands that regulate when to send and receive data.

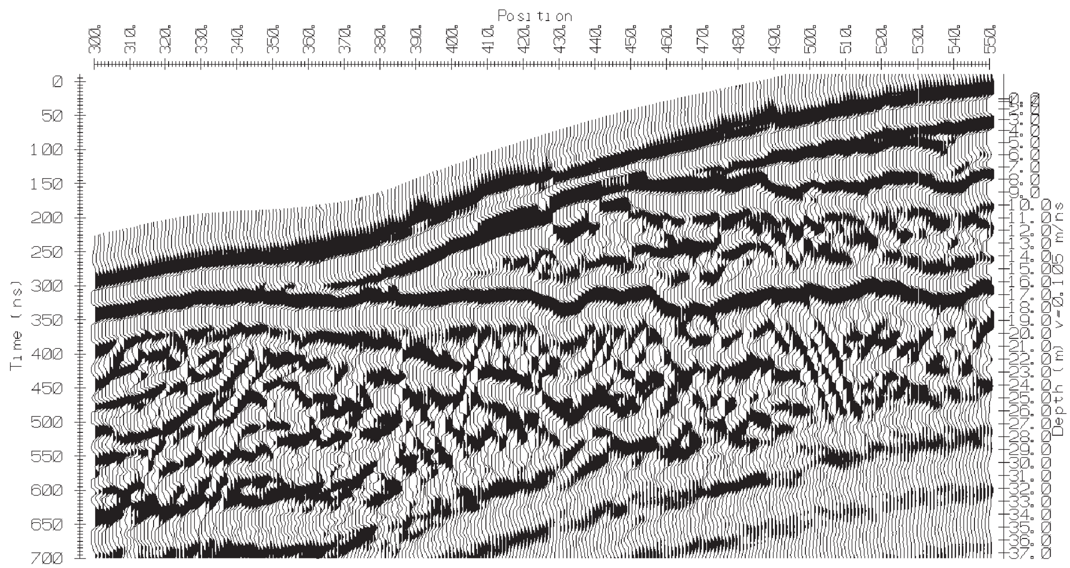


Figure 1. Topographically corrected Ground Penetrating Radar profile. Horizontal line at 350 ns time is groundwater surface.

After the data acquisition, post-processing begins. The GPS data are differentially corrected for standard errors in GPS, such as atmospheric, ephemeris (satellite positional error), and other errors that can be corrected by differential correction. These data are then converted into GIS coverages and named according to the GPR line. We refer to each set of data as a line. Each line then needs to be corrected by using heads up digitizing in the GIS software. Even though it has been differentially corrected, the GPS traverse will have "wiggles" in it due to errors in collection. These errors will make the GPS line greater in length than the GPR line and result in errors. Each line is smoothed by the GIS operator such that it matches the notes collected from the field and runs clearly down the road on which the GPR was collected. An overlay of aerial photography assists greatly in this process.

Next, a set of "addresses" is generated. Each GPR line is treated as if it were a street in the GIS software. Thus, it has a beginning "address" of 0 and an ending address of the length of the line. For example, a 1.5Km line would have a beginning address of 0 and an ending address of 1500. GIS software has a method for geocoding addresses against linear features based on the beginning and ending addresses of streets. This is often used for delivering pizzas, for example. In our case, we use this technique to determine positions along a line of data acquisition. We generate a set of addresses every 10 meters to retrieve elevation data at this interval from the LIDAR (Figure 2 shows addresses and corresponding points).

These addresses are then geocoded against the GPS lines, which generates a set of points that are 10 meters apart on the GPS line. Next, we convert these points to three-dimensional points, taking the Z-value from the LIDAR. After extracting these Z-values into a new field in the database, these values are reduced to an array of x-offset and z value (Table 1), which can be used to correct the GPR data, as seen previously in Figure 1.

CONCLUSION

This methodology has enabled us to collect data more rapidly than in the past. There are obviously some fundamental requirements that LIDAR data and aerial photography exist for the study area. With these data in place, however, we can now collect many kilometers of data per day without the expensive field collection of surveying elevation points. This has allowed us to collect and process voluminous data on groundwater surfaces in the coastal plains of Oregon. As more LIDAR data become available, we anticipate using this method to acquire better surficial deposit data for many areas.

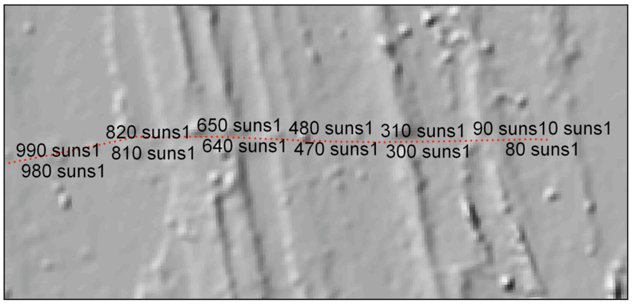


Figure 2. GPR "addresses" overlaid on a hillshade of LIDAR. Addresses are meters of offset from beginning of lines, concatenated with the name of the GPR line. North-trending linear features are shore parallel dune ridges.

Table 1. Sample elevation offsets calculated from LIDAR and geocoded GPR "addresses".

X location (m)	Z elevation (m)
0	6.8
10	7
20	7.1
30	6.3
40	5.6
50	5.1
60	5.3
70	5.4
80	5.2
90	4.9
100	4.8
110	4.8
120	4.7
130	4.8

REFERENCES

Daniels, R.C., 2001, Datum conversion issues with LIDAR spot elevation data: Photogrammetric Engineering and Remote Sensing, v.67, p.735-740.

Jol, H.M., and Bristow, C.S., 2003, GPR in sediments: advice on data collection, basic processing and interpretation, a good practice guide, *in* Bristow, C.S., and Jol, H.M., eds., GPR in Sediments: Geological Society of London, Special Publication 211, p.9-27.

Peterson, C.D., Jol, H.M., Percy, David, and Nielsen, E.L., in press, Groundwater surface trends from ground penetrating radar (GPR) profiles taken across late-Holocene barriers and beach plains of the Columbia River littoral system, Pacific Northwest coast, USA: Geological Society of America Special Publication.

GIS and GPS Utility in the Geologic Mapping of Complex Geologic Terrane on the Mascot, Tennessee 7.5' Quadrangle

By Barry W. Miller and Robert C. Price III

Tennessee Department of Environment and Conservation
Division of Geology
3711 Middlebrook Pike
Knoxville, TN 37921
Telephone: (865) 594-6200
Fax: (865) 594-6105
e-mail: Barry.Miller@state.tn.us, Bob.Price@state.tn.us

INTRODUCTION

The bedrock geology of the Mascot, Tennessee 7.5' Quadrangle was mapped in 2003 under a STATEMAP cooperative agreement between the U.S. Geological Survey and the Tennessee Division of Geology. The Mascot Quadrangle is located in the Valley and Ridge Province of east Tennessee; the bedrock consists of folded and faulted Cambrian and Ordovician strata (Figure 1). The Tennessee Division of Geology purchased two Trimble GeoExplorer 3 Global Positioning Systems (GPS), which were used in conjunction with ESRI ArcView 3.2 Geographic Information System (GIS) software to record the geologic field data and present the results of the geologic mapping.

The decision to utilize GPS technology for this mapping project was initially based on the geologic complexity of the Mascot Quadrangle and the need to improve the accuracy of geologic station location. The Division of Geology had previously used GPS units to locate oil and gas wells in a well inventory study, but the majority of well data was input into a field book or datasheet. The Mascot project would help determine how the utility of GPS and, specifically, the ability to input data directly into the GPS unit would affect the geologic mapping process.

METHODOLOGY AND RESULTS

Utilizing the Trimble GeoExplorer 3 GPS and Data Dictionary for Geologic Field Data Collection

A GPS data dictionary contains a catalog of the features and attributes pertinent to an endeavor or project. It is used in the field to control the data collection of a feature (e.g., an object, geologic station, rock outcrop, etc.) and its attributes (e.g., object information, soil type, rock lithology, etc.). Using the Trimble GPS data dictionary, a geologic data spreadsheet was created to record the important geologic aspects and their values relevant to the mapping of the Mascot Quadrangle (Figure 2). Recording geologic observations involved scrolling through attribute windows (e.g., lithology) and choosing their values (e.g., shale, limestone, sandstone) from a predefined drop-down

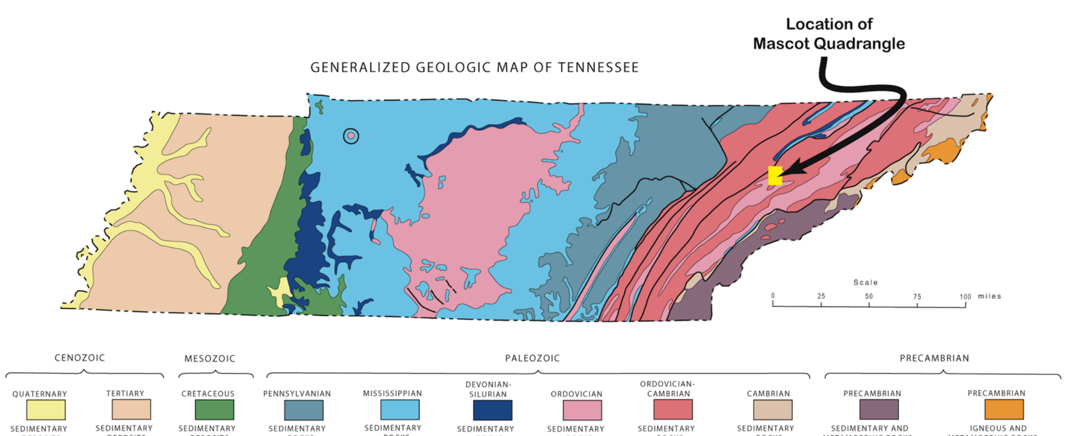
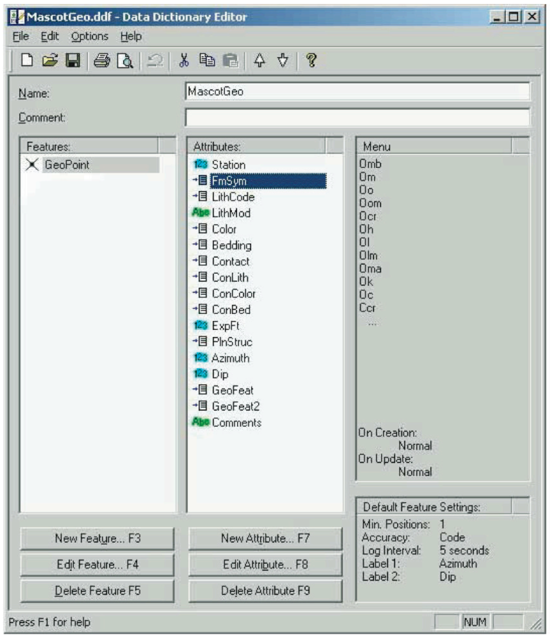


Figure 1. Location of the Mascot 7.5' Quadrangle.

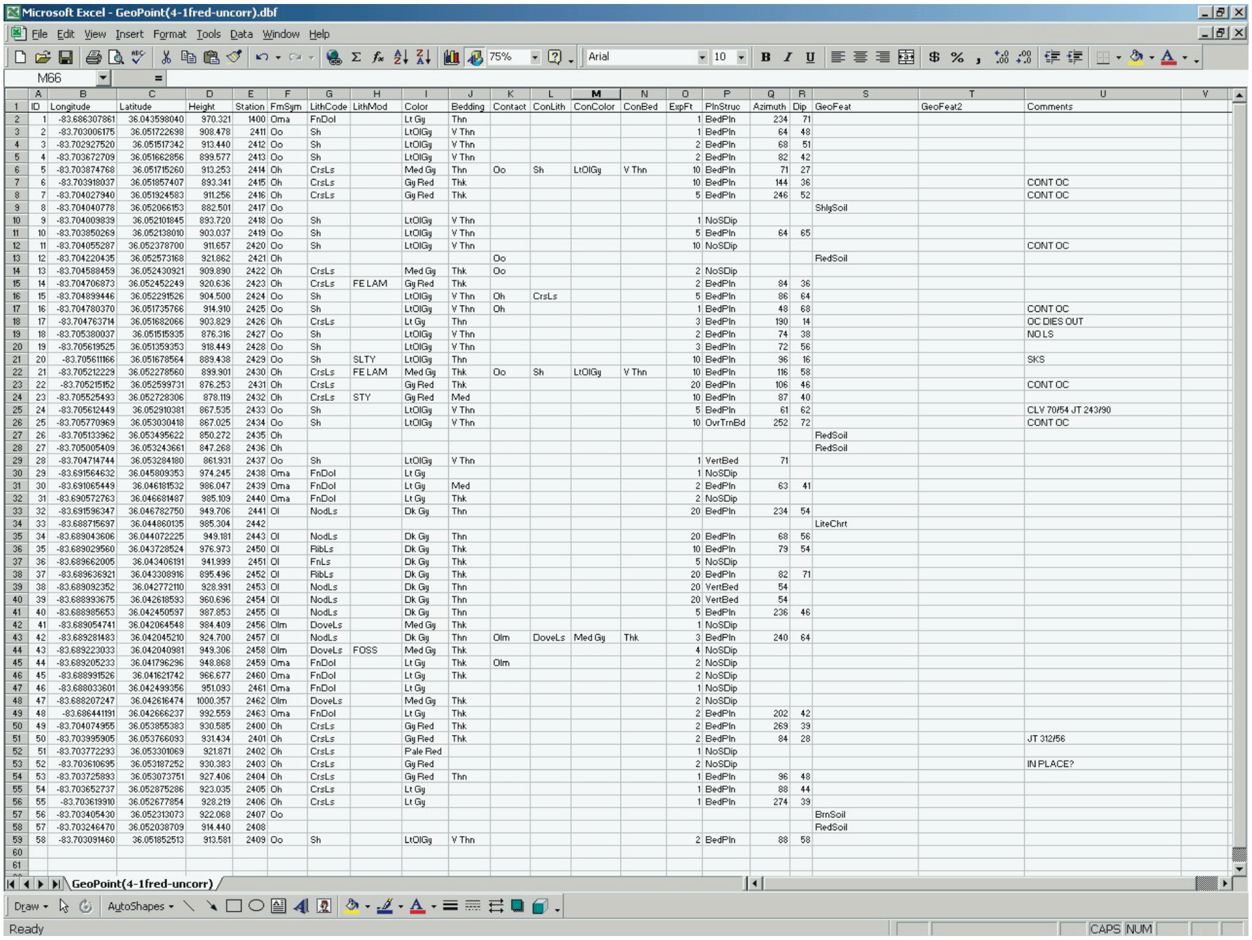


Numeric imput Field station number 1-9000	Drop-down menu Standard abbreviations of Formation names	Drop-down menu Abbreviations of diagnostic rock lithology	Text input Lithology modifiers	Drop-down menu Commonly observed lithologic colors	Drop-down menu Bedding thickness description	Drop-down menu Formation contact. Standard abbreviations of Formation names	Drop-down menu Contact lithology. Abbreviations of diagnostic rock lithology	Drop-down menu Contact color. Commonly observed lithologic colors	Drop-down menu Contact bedding. Bedding thickness description	Numeric input Top bed to bottom bed 0-300	Drop-down menu Planar structural features	Numeric input Strike in azimuth degrees 0-360	Numeric input Dip measurement 0-90	Drop-down menu Various geologic features	Drop-down menu Various geologic features repeated	Text input Other comments
Station	**FmSym**	**LithCode**	**LithMod**	**Color**	**Bedding**	**Contact**	**ConLith**	**ConColor**	**ConBed**	**ExpFt**	**PlnStruc**	**Azimuth**	**Dip**	**GeoFeat**	**GeoFeat2**	**Comments**
	Omb	Sh		Lt Gy	V Thk	Omb	Sh	Lt Gy	V Thk		BedPln			CrypChrt	CrypChrt	
	Om	SltySh		Med Gy	Thk	Om	SltySh	Med Gy	Thk		NoSDip			OoChrt	OoChrt	
	Oo	Slst		Dk Gy	Med	Oo	Slst	Dk Gy	Med		BedPln?			AlglChrt	AlglChrt	
	Oom	SdySh		LtOlGy	Thn	Oom	SdySh	LtOlGy	Thn		HorzBed			VugChrt	VugChrt	
	Ocr	Ss		Olgy	V Thn	Ocr	Ss	Olgy	V Thn		VertBed			BlkChrt	BlkChrt	
	Oho	FnLs		Grn Gy	Thk Lam	Oho	FnLs	Grn Gy	Thk Lam		OverTrnBed			LiteChrt	LiteChrt	
	Ole	CrsLs		Brn Gy	Thn Lam	Ole	CrsLs	Brn Gy	Thn Lam		Joint			RedChrt	RedChrt	
	Olm	RibLs		Rd Gy	Cross	Olm	RibLs	Rd Gy	Cross		VertJnt			PorcChrt	PorcChrt	
	Oma	NodLs		Red	? - Default	Oma	NodLs	Red	? - Default		Cleavage			MoldChrt	MoldChrt	
	Ok	DoveLs		Gy Red		Ok	DoveLs	Gy Red			VertClvg			OvalChrt	OvalChrt	
	Oc	FnDol		Pale Red		Oc	FnDol	Pale Red			Fault			AlgalBall	AlgalBall	
	Ccr	CrsDol		Gy Pink		Ccr	CrsDol	Gy Pink			Fault?			Cottonball	Cottonball	
	Cmnc	PetDol		Ylw		Cmnc	PetDol	Ylw			MnrAnt			BlkyChrt	BlkyChrt	
	Cmnl	RdClst		Ylw Gy		Cmnl	RdClst	Ylw Gy			MnrSyn			ShFloat	ShFloat	
	Cn	YlClst		Brn Gy		Cn	YlClst	Brn Gy			CrumpBed			SsFloat	SsFloat	
	Cm	Other		Red Brn		Cm	Other	Red Brn			? - Default			SsCobbls	SsCobbls	
	Crg	? - Default		Other		Crg	? - Default	Other						ChtMtxSs	ChtMtxSs	
	Crgc			? - Default		Crgc		? - Default						Breccia	Breccia	
	Crt					Crt								Hematite	Hematite	
	Cpv					Cpv								Limonite	Limonite	
	Cr					Cr								RedSoil	RedSoil	
	Crc					Crc								ShlySoil	ShlySoil	
	? - Default					? - Default								BrnSoil	BrnSoil	
														Spring	Spring	
														Depress	Depress	
														Collapse	Collapse	
														Cave	Cave	
														Opening	Opening	
														Swallet	Swallet	
														Quarry	Quarry	
														? - Default	? - Default	

Figure 2. The GPS data dictionary (top) for the Mascot 7.5' Quadrangle with expanded menu selections for the geologic attributes (bottom).

menu. Attribute menu nomenclature was abbreviated to fit the constraints of the drop-down menu screen on the GPS unit. Numerical fields for strike and dip of planar features such as bedding, joints, and cleavage required direct user input. Strikes were measured in azimuth degrees (using the right hand rule) to facilitate the proper rotation of geologic symbols when compiling the data in ArcView. Two general geologic attribute fields containing drop-down menus of additional pertinent geologic information acquired during mapping include soil character, chert type, karst features, and mining activities. A final comment field allows the mapper to input directly any other observations using a menu keypad. The data dictionary and accompanying spreadsheet were updated as needed when new useful mapping criteria were observed.

The geologic data generally were input via the data dictionary, while the GPS unit collected the satellite-based coordinate information. The latitude and longitude coordinates were collected by the GPS unit in a decimal degree format so that, later, field station locations could be plotted in ArcView (Figure 3). After returning to the office, the data files were downloaded to a computer, and the coordinates were differentially corrected to a local base station over the Internet to improve accuracy.



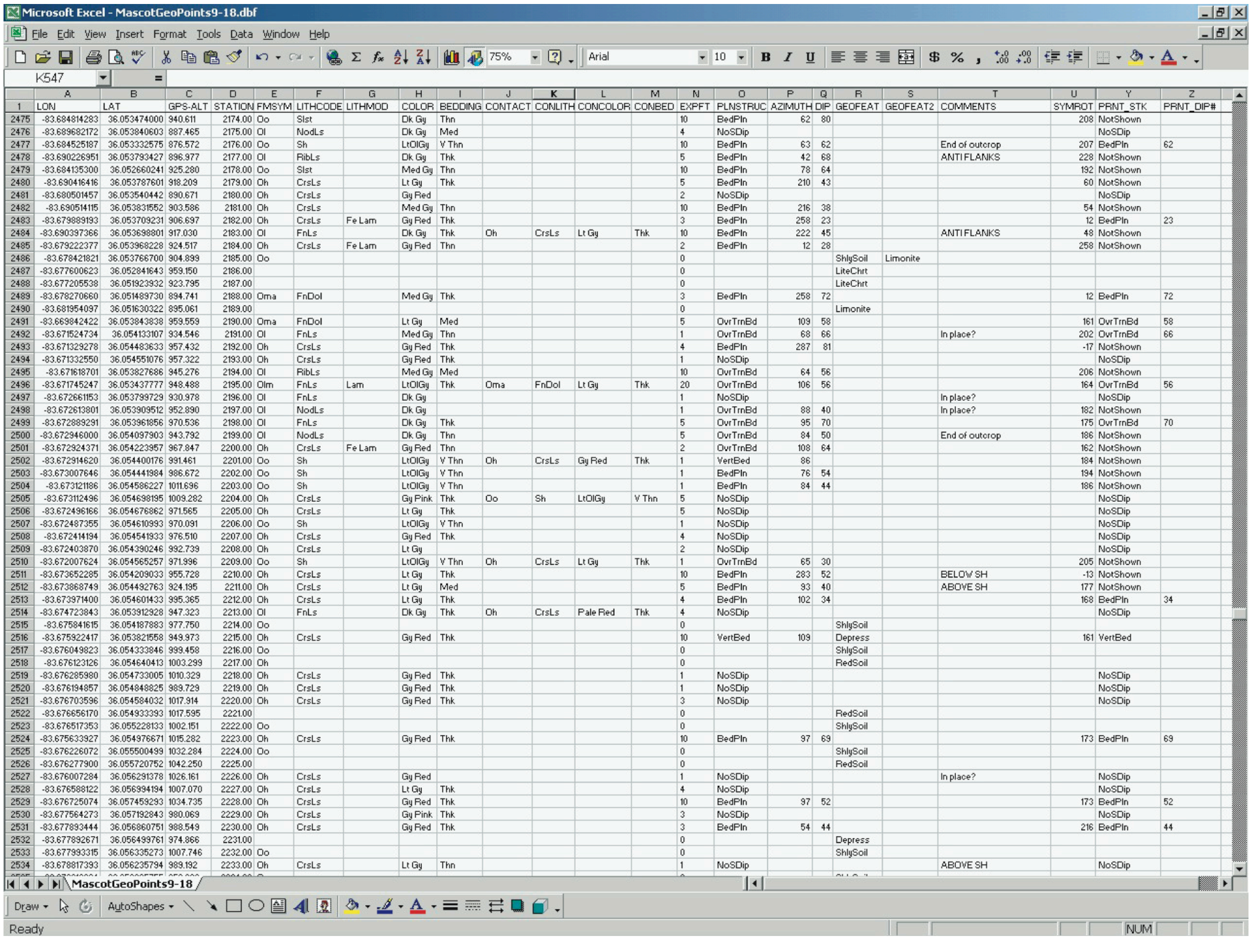
ID	Longitude	Latitude	Height	Station	FmSym	LithCode	LithMod	Color	Bedding	Contact	ConLith	ConColor	ConBed	ExpFt	PlnStruc	Azimuth	Dip	GeoFeat	GeoFeat2	Comments
1	-83.686307861	36.043598040	970.321	1400	Oma	FnDol		Lt Gy	Thn					1	BedPln	234	71			
2	-83.703006175	36.051722698	908.478	2411	Oo	Sh		LtOlGy	V Thn					1	BedPln	64	48			
3	-83.702927520	36.051517342	913.440	2412	Oo	Sh		LtOlGy	V Thn					2	BedPln	68	51			
4	-83.703672709	36.051662856	899.577	2413	Oo	Sh		LtOlGy	V Thn					2	BedPln	82	42			
5	-83.703874768	36.051715260	913.253	2414	Oh	CrsLs		Med Gy	Thn	Oo	Sh	LtOlGy	V Thn	10	BedPln	71	27			
6	-83.703918037	36.051857407	893.341	2415	Oh	CrsLs		Gy Red	Thk					10	BedPln	144	36			CONT OC
7	-83.704027940	36.051924583	911.256	2416	Oh	CrsLs		Gy Red	Thk					5	BedPln	246	52			CONT OC
8	-83.704040778	36.052066153	882.501	2417	Oo													ShlySoil		
9	-83.704009839	36.052101845	893.720	2418	Oo	Sh		LtOlGy	V Thn					1	NoSDip					
10	-83.703850269	36.052138010	903.037	2419	Oo	Sh		LtOlGy	V Thn					5	BedPln	64	65			
11	-83.704055287	36.052378700	911.657	2420	Oo	Sh		LtOlGy	V Thn					10	NoSDip					CONT OC
12	-83.704220435	36.052573168	921.862	2421	Oh					Oo								RedSoil		
13	-83.704588459	36.052430921	909.890	2422	Oh	CrsLs		Med Gy	Thk	Oo				2	NoSDip					
14	-83.704706873	36.052452249	920.636	2423	Oh	CrsLs	FE LAM	Gy Red	Thk					2	BedPln	84	36			
15	-83.704899446	36.052291526	904.500	2424	Oo	Sh		LtOlGy	V Thn	Oh	CrsLs			5	BedPln	86	64			
16	-83.704780370	36.051735766	914.910	2425	Oo	Sh		LtOlGy	V Thn	Oh				1	BedPln	48	68			CONT OC
17	-83.704763714	36.051682066	903.829	2426	Oh	CrsLs		Lt Gy	Thn					3	BedPln	190	14			OC DIES OUT
18	-83.705380037	36.051515935	876.316	2427	Oo	Sh		LtOlGy	V Thn					2	BedPln	74	38			NO LS
19	-83.705619525	36.051359353	918.449	2428	Oo	Sh		LtOlGy	V Thn					3	BedPln	72	56			
20	-83.705611166	36.051678564	889.438	2429	Oo	Sh	SLTY	LtOlGy	Thn					10	BedPln	96	16			SKS
21	-83.705212229	36.052278560	899.901	2430	Oh	CrsLs	FE LAM	Med Gy	Thk	Oo	Sh	LtOlGy	V Thn	10	BedPln	116	58			
22	-83.705215152	36.052599731	876.253	2431	Oh	CrsLs		Gy Red	Thk					20	BedPln	106	46			CONT OC
23	-83.705525493	36.052728306	878.119	2432	Oh	CrsLs	STY	Gy Red	Med					10	BedPln	87	40			
24	-83.705612449	36.052910381	867.535	2433	Oo	Sh		LtOlGy	V Thn					5	BedPln	61	62			CLV 70/54 JT 243/90
25	-83.705770969	36.053030418	867.025	2434	Oo	Sh		LtOlGy	V Thn					10	OvrTrnBd	252	72			CONT OC
26	-83.705133962	36.053495622	850.272	2435	Oh													RedSoil		
27	-83.705005409	36.053243661	847.268	2436	Oh													RedSoil		
28	-83.704714744	36.053284180	861.931	2437	Oo	Sh		LtOlGy	V Thn					1	VertBed	71				
29	-83.691564632	36.045809353	974.245	2438	Oma	FnDol		Lt Gy						1	NoSDip					
30	-83.691065449	36.046181532	986.047	2439	Oma	FnDol		Lt Gy	Med					2	BedPln	63	41			
31	-83.690572763	36.046681487	985.109	2440	Oma	FnDol		Lt Gy	Thk					2	NoSDip					
32	-83.691596347	36.046782750	949.706	2441	Ol	NodLs		Dk Gy	Thn					20	BedPln	234	54			
33	-83.688715697	36.044860135	985.304	2442														LiteChrt		
34	-83.689043606	36.044072225	949.181	2443	Ol	NodLs		Dk Gy	Thn					20	BedPln	68	56			
35	-83.689029560	36.043728524	976.973	2450	Ol	RibLs		Dk Gy	Thk					10	BedPln	79	54			
36	-83.689662005	36.043406191	941.999	2451	Ol	FnLs		Dk Gy	Thk					5	NoSDip					
37	-83.689636921	36.043308916	895.496	2452	Ol	RibLs		Dk Gy	Thk					20	BedPln	82	71			
38	-83.689092352	36.042772110	928.991	2453	Ol	NodLs		Dk Gy	Thn					20	VertBed	54				
39	-83.688993675	36.042618593	960.696	2454	Ol	NodLs		Dk Gy	Thn					20	VertBed	54				
40	-83.688985653	36.042450597	987.853	2455	Ol	NodLs		Dk Gy	Thn					5	BedPln	236	46			
41	-83.689054741	36.042064548	984.409	2456	Olm	DoveLs		Med Gy	Thk					1	NoSDip					
42	-83.689281483	36.042045210	924.700	2457	Ol	NodLs		Dk Gy	Thn	Olm	DoveLs	Med Gy	Thk	3	BedPln	240	64			
43	-83.689223033	36.042040981	949.306	2458	Olm	DoveLs	FOSS	Med Gy	Thk					4	NoSDip					
44	-83.689205233	36.041796296	948.868	2459	Oma	FnDol		Lt Gy	Thk	Olm				2	NoSDip					
45	-83.688991526	36.041621742	966.677	2460	Oma	FnDol		Lt Gy	Thk					2	NoSDip					
46	-83.688033601	36.042499356	951.093	2461	Oma	FnDol		Lt Gy						1	NoSDip					
47	-83.688207247	36.042616474	1000.357	2462	Olm	DoveLs		Med Gy	Thk					2	NoSDip					
48	-83.686441191	36.042666237	992.559	2463	Oma	FnDol		Lt Gy	Thk					2	BedPln	202	42			
49	-83.704074955	36.053855383	930.585	2400	Oh	CrsLs		Gy Red	Thk					2	BedPln	269	39			
50	-83.703995905	36.053766093	931.434	2401	Oh	CrsLs		Gy Red	Thk					2	BedPln	84	28			JT 312/56
51	-83.703772293	36.053301069	921.871	2402	Oh	CrsLs		Pale Red						1	NoSDip					
52	-83.703610695	36.053187252	930.383	2403	Oh	CrsLs		Gy Red						2	NoSDip					IN PLACE?
53	-83.703725893	36.053073751	927.406	2404	Oh	CrsLs		Gy Red	Thn					1	BedPln	96	48			
54	-83.703652737	36.052875286	923.035	2405	Oh	CrsLs		Lt Gy						1	BedPln	88	44			
55	-83.703619910	36.052677854	928.219	2406	Oh	CrsLs		Lt Gy						1	BedPln	274	39			
56	-83.703405430	36.052313073	922.068	2407	Oo													BrnSoil		
57	-83.703246470	36.052038709	914.440	2408														RedSoil		
58	-83.703091460	36.051852513	913.581	2409	Oo	Sh		LtOlGy	V Thn					2	BedPln	88	58			

Figure 3. Example of part of the fieldwork database for the Mascot 7.5' Quadrangle.

Utilizing ESRI ArcView 3.2 GIS Software and the GPS Database to Produce the Mascot Quadrangle Geologic Map

The differentially corrected GPS data file (.cor) was exported in dBase format and transferred into a compilation database that was used in ArcView to compile a preliminary geologic map. Additional fields in the compilation database are included to plot the symbol orientation for planar features (SYMROT) and print selected geologic symbols (PRNT_STK) and dip numbers (PRNT_DIP#) in ArcView (Figure 4). The SYMROT field was used to rotate the geologic symbols by using the formula "270° minus the azimuth strike." The compilation map included point themes for lithology, bedding attitudes, formation contact points, and other geologic features recorded in the database, all of which are shown as unique points or symbols on the map (Figure 5 and 6). Dip value labels were added to the bedding attitude symbols. These point themes were used to interpret the location of stratigraphic contacts and the surface trace of axial planes and faults, which are each separate line themes on the geologic compilation map (Figure 7).

To produce the "final" version of the geologic map in ArcView, stratigraphic contacts and fault lines were divided into solid (exactlocation), dashed (approximatelocation), and dotted (coveredlocation) line segments based on contact location certainty. These contact line themes were converted to formation polygon themes. The formation polygons were colored and displayed beneath a partially transparent raster topographic base map. The color pallet of the raster map was adjusted such that the underlying polygons could be displayed while important features such as roads, streams, and contour lines also were visible. Formation labels then were added to the geologic map. To avoid clutter, the lithologic and geologic feature point themes were not displayed in the final version of the geologic map (Figure 8). A geologic cross-section location line was placed on the map. In the ArcView layout view, titles, labels, and a geologic symbol explanation were added as well as a written scale, bar scales, and a north arrow (Figure 9). The geologic cross-section, stratigraphic column, and geologic descriptions were drafted separately using Adobe Illustrator 8.0 software.



1	LON	LAT	GPS-ALT	STATION	FMSYM	LITHCODE	LITHMOD	COLOR	BEDDING	CONTACT	CONLITH	CONCOLOR	CONBED	EXPFT	PLNSTRUC	AZIMUTH	DIP	GEOFEAT	GEOFEAT2	COMMENTS	SYMROT	PRNT_STK	PRNT_DIP#
2475	-83.684814283	36.053474000	940.611	2174.00	Oo	Slst		Dk Gy	Thn					10	BedPln	62	80				208	NotShown	
2476	-83.689682172	36.053840603	887.465	2175.00	Ol	NodLs		Dk Gy	Med					4	NoSDip							NoSDip	
2477	-83.684525187	36.053332575	876.572	2176.00	Oo	Sh		LtOlGy	V Thn					10	BedPln	63	62			End of outcrop	207	BedPln	62
2478	-83.690226951	36.053793427	896.977	2177.00	Ol	RibLs		Dk Gy	Thk					5	BedPln	42	68			ANTI FLANKS	228	NotShown	
2479	-83.684135300	36.052660241	925.280	2178.00	Oo	Slst		Med Gy	Thn					10	BedPln	78	64				192	NotShown	
2480	-83.690416416	36.053787601	918.209	2179.00	Oh	CrsLs		Lt Gy	Thk					5	BedPln	210	43				60	NotShown	
2481	-83.680501457	36.053540442	890.671	2180.00	Oh	CrsLs		Gy Red						2	NoSDip							NoSDip	
2482	-83.690514115	36.053831552	903.586	2181.00	Oh	CrsLs		Med Gy	Thn					10	BedPln	216	38				54	NotShown	
2483	-83.679889193	36.053709231	906.697	2182.00	Oh	CrsLs	Fe Lam	Gy Red	Thk					3	BedPln	258	23				12	BedPln	23
2484	-83.690397366	36.053698801	917.030	2183.00	Ol	FnLs		Dk Gy	Thk	Oh	CrsLs	Lt Gy	Thk	10	BedPln	222	45			ANTI FLANKS	48	NotShown	
2485	-83.679222377	36.053968228	924.517	2184.00	Oh	CrsLs	Fe Lam	Gy Red	Thn					2	BedPln	12	28				258	NotShown	
2486	-83.678421821	36.053766700	904.899	2185.00	Oo									0				ShlySoil	Limonite				
2487	-83.677600623	36.052841643	959.150	2186.00										0				LiteChrt					
2488	-83.677205538	36.051923932	923.795	2187.00										0				LiteChrt					
2489	-83.678270660	36.051489730	894.741	2188.00	Oma	FnDol		Med Gy	Thk					3	BedPln	258	72				12	BedPln	72
2490	-83.681954097	36.051630322	895.061	2189.00										0				Limonite					
2491	-83.669842422	36.053843838	959.559	2190.00	Oma	FnDol		Lt Gy	Med					5	OvrTrnBd	109	58				161	OvrTrnBd	58
2492	-83.671524734	36.054133107	934.546	2191.00	Ol	FnLs		Med Gy	Thn					1	OvrTrnBd	68	66			In place?	202	OvrTrnBd	66
2493	-83.671329278	36.054483633	957.432	2192.00	Oh	CrsLs		Gy Red	Thk					4	BedPln	287	81				-17	NotShown	
2494	-83.671332550	36.054551076	957.322	2193.00	Oh	CrsLs		Gy Red	Thk					1	NoSDip							NoSDip	
2495	-83.671618701	36.053827686	945.276	2194.00	Ol	RibLs		Med Gy	Med					10	OvrTrnBd	64	56				206	NotShown	
2496	-83.671745247	36.053437777	948.488	2195.00	Olm	FnLs	Lam	LtOlGy	Thk	Oma	FnDol	Lt Gy	Thk	20	OvrTrnBd	106	56				164	OvrTrnBd	56
2497	-83.672661153	36.053799729	930.978	2196.00	Ol	FnLs		Dk Gy						1	NoSDip					In place?		NoSDip	
2498	-83.672613801	36.053909512	952.890	2197.00	Ol	NodLs		Dk Gy						1	OvrTrnBd	88	40			In place?	182	NotShown	
2499	-83.672889291	36.053961856	970.536	2198.00	Ol	FnLs		Dk Gy	Thk					5	OvrTrnBd	95	70				175	OvrTrnBd	70
2500	-83.672946000	36.054097903	943.792	2199.00	Ol	NodLs		Dk Gy	Thn					5	OvrTrnBd	84	50			End of outcrop	186	NotShown	
2501	-83.672924371	36.054223957	967.847	2200.00	Oh	CrsLs	Fe Lam	Gy Red	Thn					2	OvrTrnBd	108	64				162	NotShown	
2502	-83.672914620	36.054400176	991.461	2201.00	Oo	Sh		LtOlGy	V Thn	Oh	CrsLs	Gy Red	Thk	1	VertBed	86					184	NotShown	
2503	-83.673007646	36.054441984	986.672	2202.00	Oo	Sh		LtOlGy	V Thn					1	BedPln	76	54				194	NotShown	
2504	-83.673121186	36.054586227	1011.696	2203.00	Oo	Sh		LtOlGy	V Thn					1	BedPln	84	44				186	NotShown	
2505	-83.673112496	36.054698195	1009.282	2204.00	Oh	CrsLs		Gy Pink	Thk	Oo	Sh	LtOlGy	V Thn	5	NoSDip							NoSDip	
2506	-83.672496166	36.054676862	971.565	2205.00	Oh	CrsLs		Lt Gy	Thk					5	NoSDip							NoSDip	
2507	-83.672487355	36.054610993	970.091	2206.00	Oo	Sh		LtOlGy	V Thn					1	NoSDip							NoSDip	
2508	-83.672414194	36.054541933	976.510	2207.00	Oh	CrsLs		Gy Red	Thk					4	NoSDip							NoSDip	
2509	-83.672403870	36.054390246	992.739	2208.00	Oh	CrsLs		Lt Gy						2	NoSDip							NoSDip	
2510	-83.672007624	36.054565257	971.996	2209.00	Oo	Sh		LtOlGy	V Thn	Oh	CrsLs	Lt Gy	Thk	1	OvrTrnBd	65	30				205	NotShown	
2511	-83.673652285	36.054209033	955.728	2210.00	Oh	CrsLs		Lt Gy	Thk					10	BedPln	283	52			BELOW SH	-13	NotShown	
2512	-83.673868749	36.054492763	924.195	2211.00	Oh	CrsLs		Lt Gy	Med					5	BedPln	93	40			ABOVE SH	177	NotShown	
2513	-83.673971400	36.054601433	995.365	2212.00	Oh	CrsLs		Lt Gy	Thk					4	BedPln	102	34				168	BedPln	34
2514	-83.674723843	36.053912928	947.323	2213.00	Ol	FnLs		Dk Gy	Thk	Oh	CrsLs	Pale Red	Thk	4	NoSDip							NoSDip	
2515	-83.675841615	36.054187883	977.750	2214.00	Oo									0				ShlySoil					
2516	-83.675922417	36.053821558	949.973	2215.00	Oh	CrsLs		Gy Red	Thk					10	VertBed	109		Depress			161	VertBed	
2517	-83.676049823	36.054333846	999.458	2216.00	Oo									0				ShlySoil					
2518	-83.676123126	36.054640413	1003.299	2217.00	Oh									0				RedSoil					
2519	-83.676285980	36.054733005	1010.329	2218.00	Oh	CrsLs		Gy Red	Thk					1	NoSDip							NoSDip	
2520	-83.676194857	36.054848825	989.729	2219.00	Oh	CrsLs		Gy Red	Thk					1	NoSDip							NoSDip	
2521	-83.676703596	36.054584032	1017.914	2220.00	Oh	CrsLs		Gy Red	Thk					3	NoSDip							NoSDip	
2522	-83.676656170	36.054933393	1017.595	2221.00										0				RedSoil					
2523	-83.676517353	36.055228133	1002.151	2222.00	Oo									0				ShlySoil					
2524	-83.675633927	36.054976671	1015.282	2223.00	Oh	CrsLs		Gy Red	Thk					10	BedPln	97	69				173	BedPln	69
2525	-83.676226072	36.055500499	1032.284	2224.00	Oo									0				ShlySoil					
2526	-83.676277900	36.055720752	1042.250	2225.00										0				RedSoil					
2527	-83.676007284	36.056291378	1026.161	2226.00	Oh	CrsLs		Gy Red						1	NoSDip					In place?		NoSDip	
2528	-83.676588122	36.056994194	1007.070	2227.00	Oh	CrsLs		Lt Gy	Thk					4	NoSDip							NoSDip	
2529	-83.676725074	36.057459293	1034.735	2228.00	Oh	CrsLs		Gy Red	Thk					10	BedPln	97	52				173	BedPln	52
2530	-83.677564273	36.057192843	980.069	2229.00	Oh	CrsLs		Gy Pink	Thk					3	NoSDip							NoSDip	
2531	-83.677893444	36.056860751	988.549	2230.00	Oh	CrsLs		Gy Red	Thk					3	BedPln	54	44				216	BedPln	44
2532	-83.677892671	36.056499761	974.866	2231.00										0				Depress					
2533	-83.677993315	36.056335273	1007.746	2232.00	Oo									0				ShlySoil					
2534	-83.678817393	36.056235794	989.192	2233.00	Oh	CrsLs		Lt Gy	Thn					1	NoSDip					ABOVE SH		NoSDip	

Figure 4. Example of part of the compilation database for the Mascot 7.5' Quadrangle. Note SYMROT, PRNT_STK, and PRNT_DIP columns near the right side of figure.

ADVANTAGES AND DISADVANTAGES ENCOUNTERED

Advantages of using GIS and GPS technology during the geologic mapping process included the drastic reduction of office time to input field data into a spreadsheet format that can be used in ArcView. Prior to using GPS, office time was required to determine field point coordinates and enter field book data into a spreadsheet. Also, the GPS data dictionary facilitated the compilation of data from two mappers in that it listed basic attributes to record and constrained how the attributes were described. The GPS dictionary also provided a framework by which to set up a consistent GIS compilation database between separate mappers, and share data between mappers on a daily basis.

Disadvantages included the need for a paper copy of the topographic map in the field for navigation, station point location, and plotting contact lines. The paper map was also needed to plot the data in the field to determine where the geologic traverse needed to proceed. Areas that contained observable geologic characteristics, but were deemed inaccessible, were also plotted on the paper map. Field books were used for lengthy geologic descriptions, sketches, or other information. GPS satellite acquisition was occasionally hampered under heavy forest canopy or when satellites were not in proper array. On these occasions, station locations were plotted directly on the paper map, and the correct coordinates were added later to the compilation database. There were no independent checks and balances in place designed to detect GPS user input errors. The ArcView compilation map was also printed on a periodic basis to guide future traverses and provide a map for the geologists to use in the field.

This methodology may not be suitable when geologic mapping involves long stays in the wilderness away from the office and electricity. In this study, data files needed to be downloaded every day or two because of GPS memory limitations and the GPS batteries had to be recharged every day.

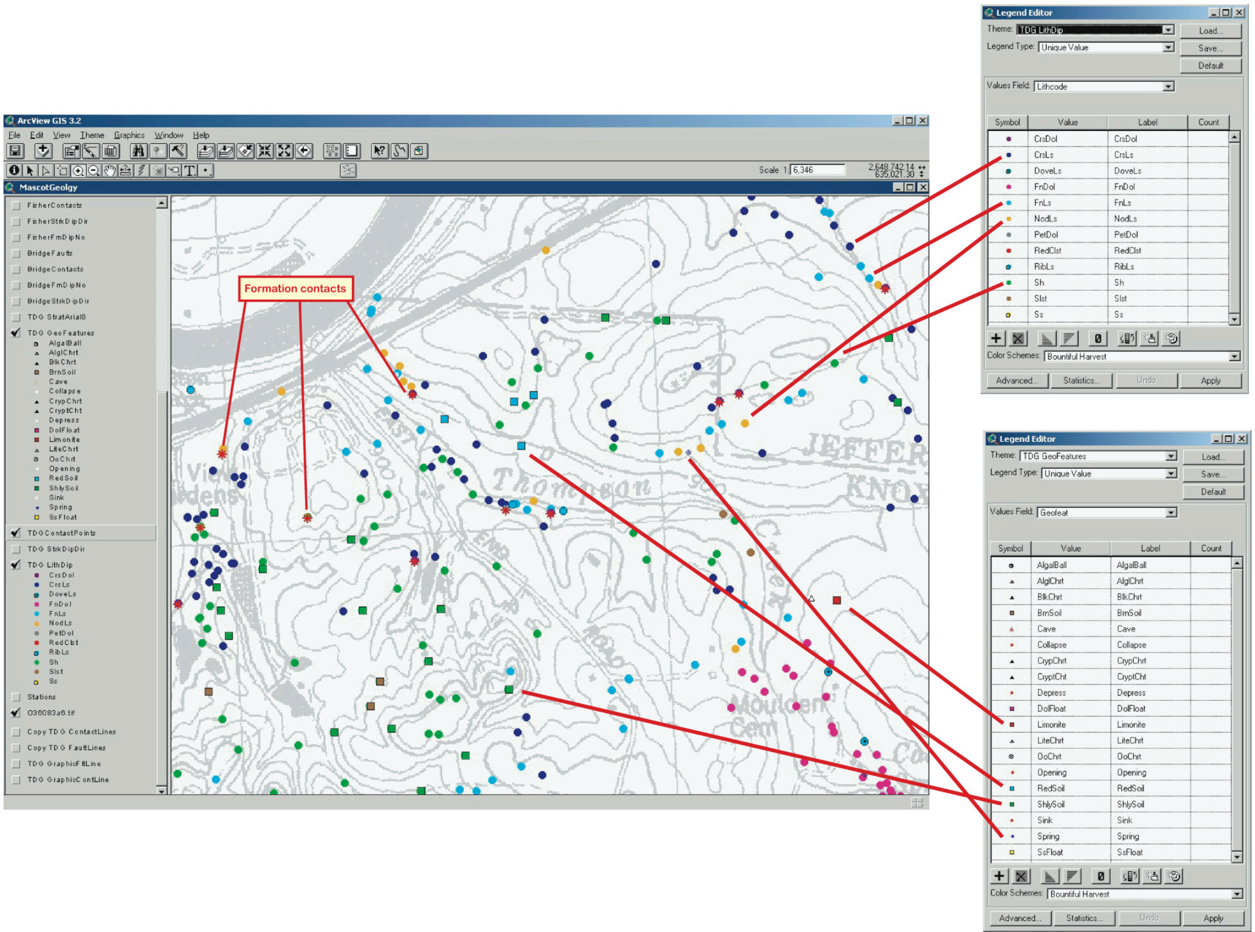


Figure 5. Part of the Mascot 7.5' Quadrangle compilation map with lithologies, geologic contacts, and other geologic features shown as discrete points or symbols.

CONCLUSIONS

Despite minor problems in the field and office, the Trimble GeoExplorer 3 Global Positioning System and the ESRI ArcView 3.2 Geographic Information System worked well in unison to assist in the completion of the geologic map of the Mascot Quadrangle. The GPS unit was used to locate station points in geologically complex areas where accurate plotting was crucial for constraining the geologic interpretation. The GPS data dictionary permitted relatively rapid data entry into the database, while the GPS unit collected coordinate information. The GIS software had the versatility to import the GPS data files directly for rapid compilation of the geologic data into a useable map. The Mascot 7.5' Geologic Quadrangle map can be printed on demand and is currently available from the Tennessee Division of Geology as an open file map.

VENDOR CONTACT INFORMATION

Adobe Illustrator—Adobe Systems, Inc., 345 Park Ave., San Jose, CA 95110-2704 USA, (800) 833-6687, <http://www.adobe.com>.

ArcGIS, ArcPad, ArcView—Environmental Systems Research Institute (ESRI), Inc., 380 New York St., Redlands, CA, 92373-8100 USA, (909) 793-2853, <http://www.esri.com>.

dBase—dataBased Intelligence, Inc., 2548 Vestal Parkway, East Vestal, NY 13850, (877) 322-7340, <http://www.dbase.com>.

Microsoft Excel—Microsoft Corp., One Microsoft Way, Redmond, WA 98052-6399 USA, (425) 882-8080, <http://www.microsoft.com/office/excel>.

Trimble GeoExplorer—Trimble Navigation Limited, 645 N. Mary Avenue, Sunnyvale, CA, 94088-3642, (408) 481-8000, <http://www.trimble.com>.

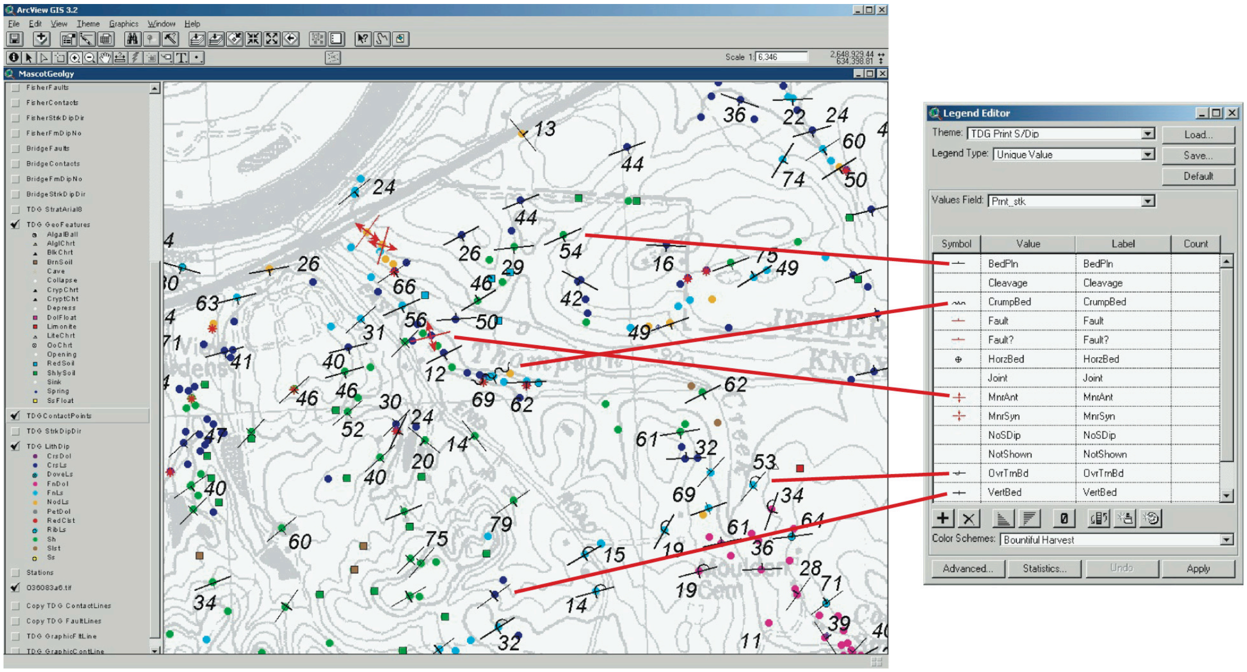


Figure 6. Part of the Mascot 7.5’ Quadrangle compilation map with bedding strike and dip symbols for the point features shown in Figure 5.

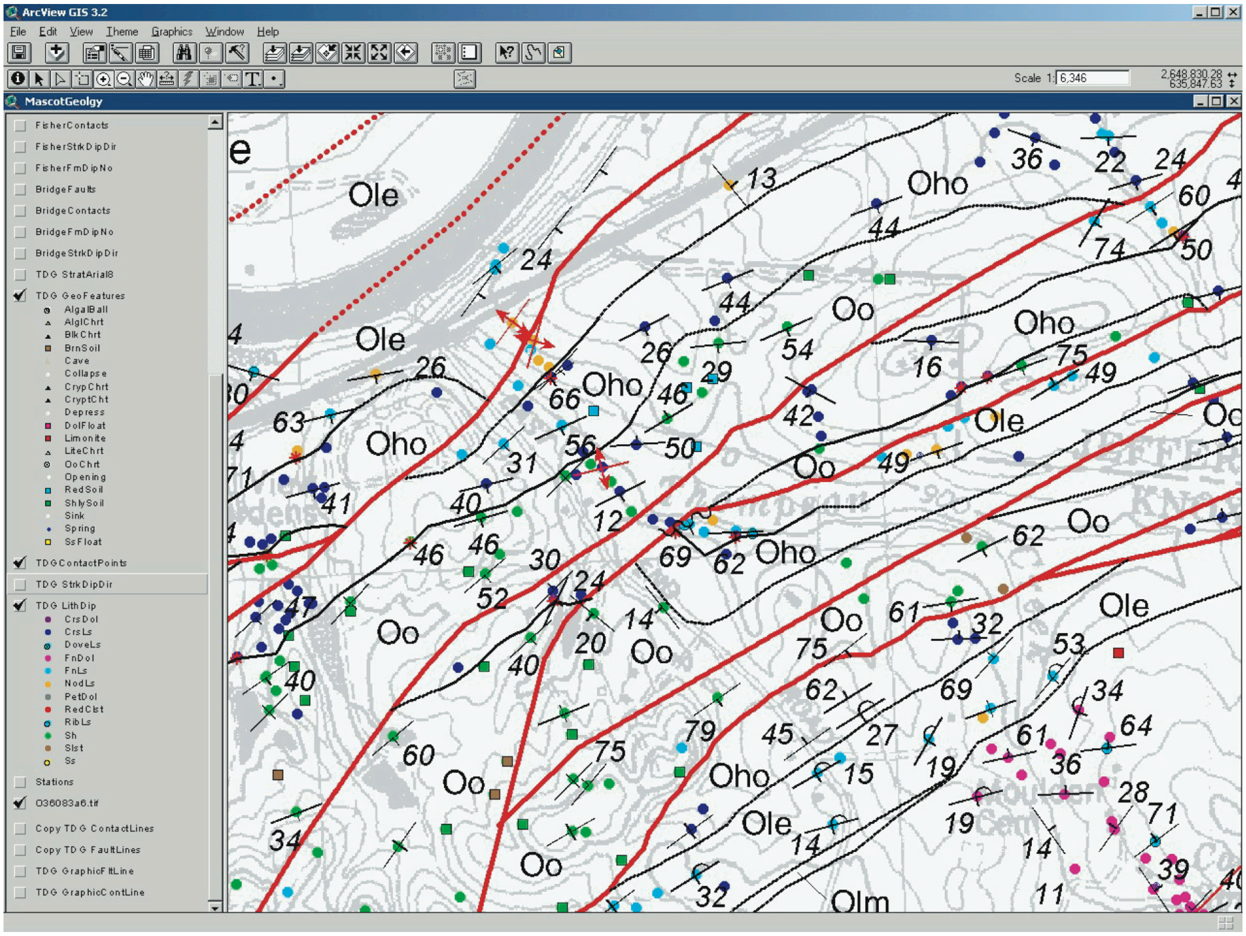


Figure 7. Part of the Mascot 7.5’ Quadrangle compilation map with geologic contact lines (black), fault lines (red), and formation abbreviations. In the printed version, the fault lines appear as thick black lines. Map includes geologic point features (see Figure 5) and bedding attitude symbology (see Figure 6).

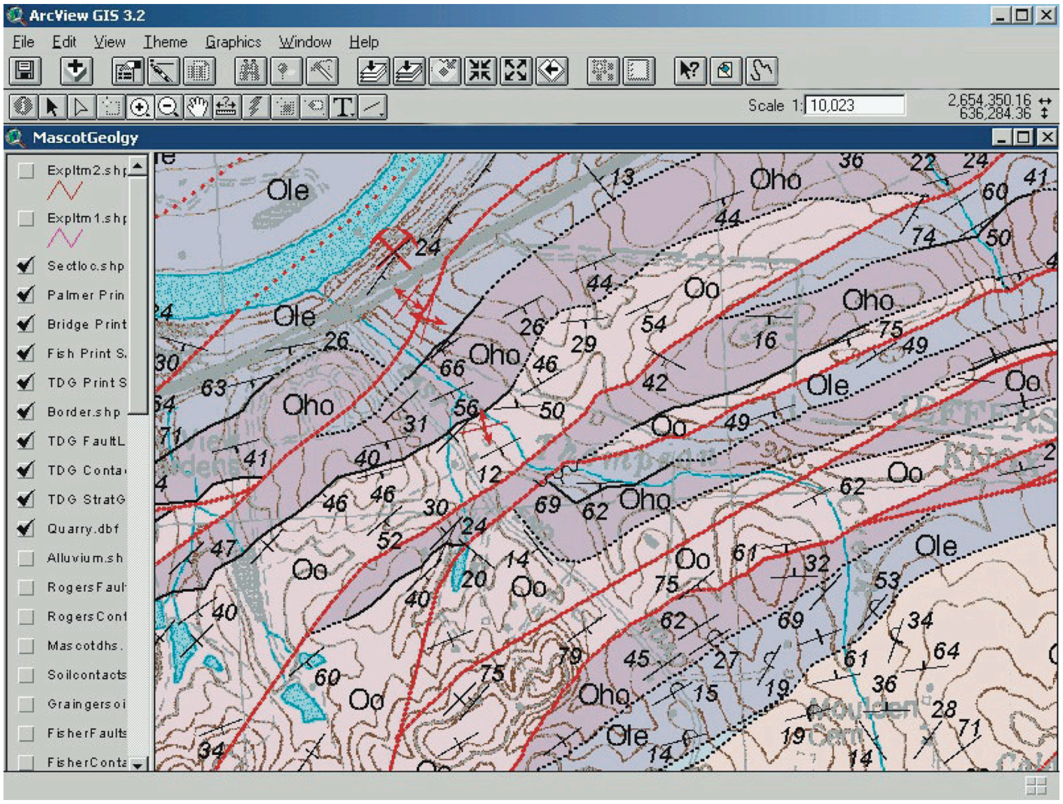


Figure 8. Part of the Mascot 7.5' Quadrangle compilation map with polygons filled with appropriate geologic formation colors. Geologic point features are not shown. Bedding attitude symbology shown.

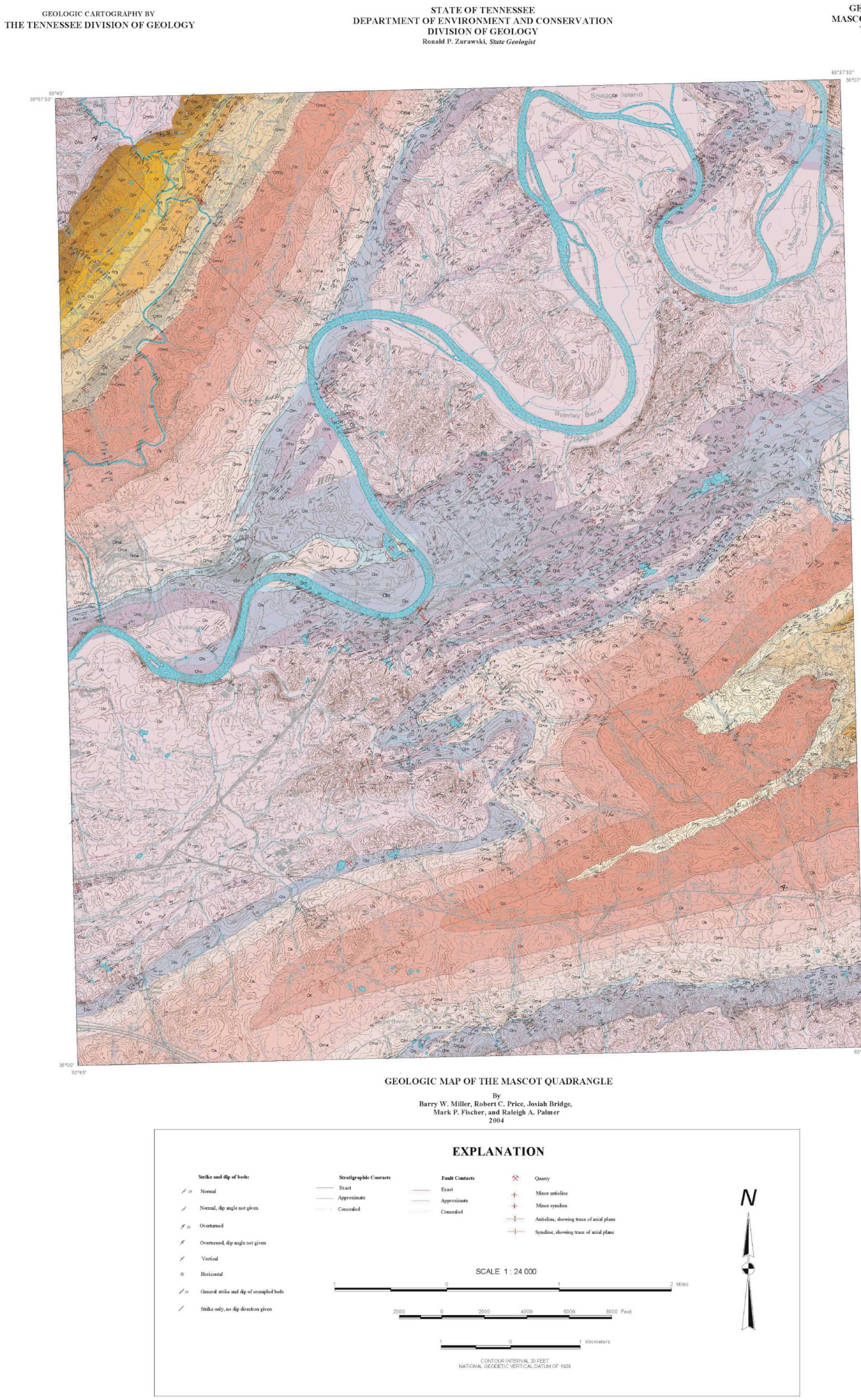


Figure 9. Open file version of the Mascot Geologic Quadrangle Map.

Prototype GIS Database for the DNAG Geologic Map of North America

By Christopher P. Garrity and David R. Soller

U.S. Geological Survey
National Center, MS 956
12201 Sunrise Valley Drive
Reston, VA 20192
Telephone: (703) 648-6426
Fax: (703) 648-6977
e-mail: {cgarrity, drsoller}@usgs.gov

PURPOSE[1]

When plans for the Geologic Map of North America (GMNA) were being made, the notion of geologic map databases was in its infancy. At that time, and for many years thereafter, few geologists were familiar with the design and use of databases to manage geologic map information. In 1998, the Geological Society of America (GSA) and the United States Geological Survey (USGS) National Geologic Map Database project agreed to cost-share the digital preparation of this map. The plan was to digitize the hand-drawn, author-prepared geologic compilations for the four map quadrants to provide digital data for two purposes: (1) to allow GSA to print the map, and (2) to permit the National Geologic Map Database project to develop a prototype database for this map. With the map now printed, the National Geologic Map Database project has begun to design and create the prototype, based on certain assumptions regarding the anticipated content of, and uses for, the map database. The first version of this database will contain the descriptive information for geologic units shown on the map. It will serve as the fundamental entity from which products of the map can be derived. These products may be interpretive, or they may be future editions of the map.

In mid-2006, this prototype will be provided to the organizations principally responsible for map compilation (GSA, USGS, GSC, and WHOI) to initiate discussion and decisions on how the map database will be designed, managed, and distributed served to the public and cooperators. The prototype is shown here to generate technical discussion and guidance prior to formal discussion among those organizations.

To produce any future editions of the map, the database will incorporate all map revisions that are necessitated by detection of compilation errors and by new regional mapping and interpretations. Further, the geologic unit descriptions shown on the printed map can be supplemented in the database by more detailed, richly attributed information derived from the many sources that were used to compile the map. This capability to revise the printed map and include additional descriptive information for map units is one of the primary reasons for building the database; the other reason is, of course, the analytical capabilities made possible by providing the map in a digital, Geographic Information System (GIS) compatible format.

The creation of this database and its enhancement to include new mapping and more richly attributed information is a daunting task that will take a significant amount of time and effort. Recognizing that a group of dedicated and knowledgeable scientists is essential to make this database useful and to keep its content up to date, GSA will develop a consortium of geological agencies to manage the database. With prototype development of the database, the National Geologic Map Database project provides a basis for this consortium to proceed.

PROTOTYPE AREA

The GMNA prototype spans an area of about 530,206 square miles and includes both continental and seafloor geologic units in the United States and Canada (Figure 1). The area was chosen due to its relatively complex geology and abundance of both onshore and offshore map symbology displayed in the published GMNA. The prototype contains over 2,500 individual polygon features and about 5,700 line features. Polygon features are symbolized by 205 unique geologic unit values, each with custom color and pattern fills designed to mimic those used in the published GMNA (Figure 2). The prototype contains customized line symbology which nearly duplicates the symbol sets used in the published map. All custom symbol sets are stored in ArcGIS layer files. Line features are classified by

[1]Modified from Soller, in Reed et al. (2005)

Figure 1. The 2005 Geologic Map of North America (GMNA) is the first such map published in the past four decades. The map, which covers 15% of the Earth’s surface (shown in yellow box), depicts the geology of the seafloor in detail never before seen on a map of this scale (1:5,000,000). It is the first geologic map of North America to be compiled since the general acceptance of plate-tectonic theory and since radiometric dates for plutonic and volcanic rocks because widely available. This map distinguishes more than 900 rock unites, 100 of which are offshore. It depicts more than seven times as many terrestrial units as are shown on the previous 1965 map, as well as detailed features of the seafloor, such as spreading centers, seamont chains, and subduction zones (Reed, et. al., 2005). The GIS prototype area (shown in red box) was chosen due to its relatively complex geology and abundance of both on- and offshore map symbology displayed in the published GMNA. The release of the prototype is meant to serve as a forum for both general comments on the overall objectives of the GIS database design, and specific comment son elements such as cartographic symbolization style.

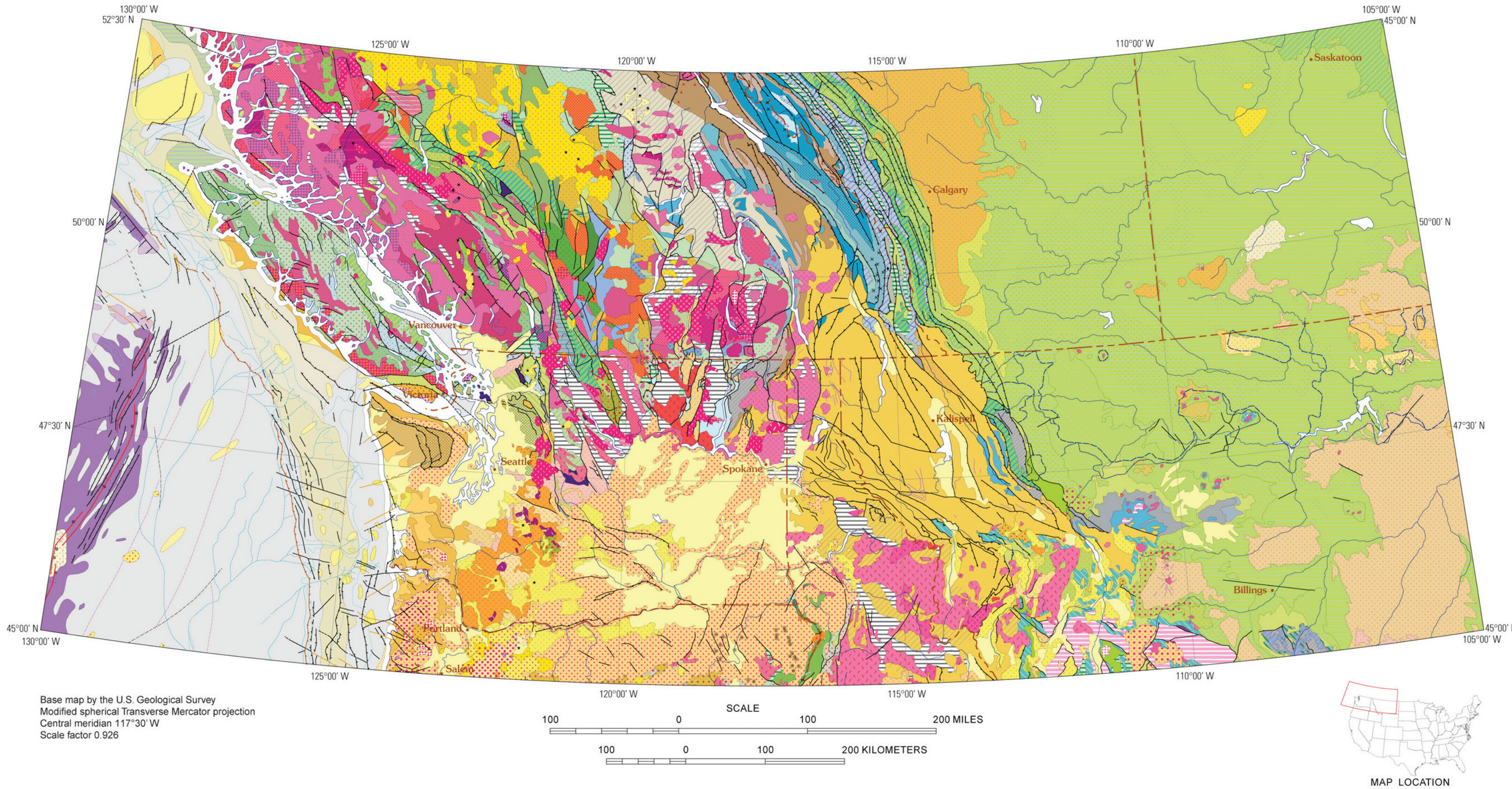


Figure 2. Layout of the prototype area exported from ArcMap. The prototype, spanning an area of about 530,206 square miles, contains over 2,500 individual polygon features and about 5,700 line features. Polygon features are symbolized by 205 unique geologic unit values, each with custom color and pattern fills designed to mimic those used in the published GMNA.

geologic line type into 26 subtypes. Geologic line types include general geologic features (contacts, faults, etc.), special submarine features (slump scars, seamount chains, spreading centers, etc.), and lithologic/age defined dikes and sills. Each subtype definition matches the explanation of the corresponding map symbol used in the GMNA.

PROCESS DESCRIPTION

A secondary objective in creation of the prototype was to determine the most efficient way to convert such an enormous digital map into a useable GIS. It was also important to determine a realistic time frame in which a project of this size could be completed. The two Adobe Illustrator files that contain the source digital data were massive, with layer counts totaling over 1,500. When analyzed at scales much greater than the intended map scale, the files showed areas where problems in topological relationships existed. Common topological problems in these areas included polygons that overlapped or had gaps between them, overlying line layers (contacts, faults, etc.) that were not coincident with polygon boundaries, and line features that self-overlapped. Although these areas were much too small to compromise the quality of the hard copy layouts, they did present problems when validating the topology in a GIS. Due to the sheer number of imported features, the errors reported after validating topology numbered in the tens of thousands. To avoid the time consuming process of correcting each error, it was decided that only the non-contact linework coincident with geologic unit boundaries would be imported directly from Adobe Illustrator. For attribution purposes, each line type (inferred thrust fault, concealed thrust fault, etc.) was imported to the GIS individually. The remaining linework (contact layer) was isolated in Adobe Illustrator and exported as a high-resolution raster image. The image was georeferenced using control points in the DNAG projection (Snyder, 1987), and auto-vectorized in ArcScan. By setting topology rules in ArcMap, line dangles in the vectorized layer were snapped to the nearest unit-bordering line feature, which resulted in a topologically clean layer. To build unit attributes quickly, individual geologic unit layers were batch exported from Adobe Illustrator and used to overlay the newly created layer. Through spatial querying, polygons in the new layer that had their center within a specific overlay layer were attributed based on the overlay’s geologic unit abbreviation. Much of the remaining attribution was completed through simple VBA field calculator scripting based on the unit abbreviation field. The final step, which proved to be the most time consuming, involved recreation of feature class symbolization in ArcGIS. 26 symbolized line types, and 205 unique color fills/patterns had to be created to mimic those in the source Adobe Illustrator file. The prototype was completed in about 10 days. GIS compatible files for the southern map sheet, which comprises the conterminous United States, are scheduled to be completed and released in FY2007. Release of GIS compatible files for the northern map sheet is scheduled to follow.

PROTOTYPE DATABASE

This prototype is intended to serve as a forum for comments on the overall objectives of the GIS database and the attribute information within. It is certain that as the database evolves, attribute information will be modified to make the database more useful. For the prototype database, a preliminary set of attributes was chosen to serve as a foundation for an eventual GMNA data model. When that data model is formalized, we anticipate that it will incorporate elements of the North American Data Model (http://nadm-geo.org/) and the International Geological Map of Europe’s data model (http://www.bgr.de/karten/IGME5000/igme5000.htm). The selected attributes attempt to capture the information depicted in the GMNA explanation sheet that accompanies the published map. The attribute list includes:

- ROCKTYPE – the “top level” rock classification (sedimentary, plutonic, volcanic, metamorphic)
- LITHOLOGY – the simplified description included for each geologic unit on the explanation sheet of the GMNA
- ROCK_UNIT_NOTE – special notes associated with certain units on the explanation sheet of the GMNA. For example, selected volcanic rocks are attributed “Basalt adjacent to active spreading centers,” selected metamorphic rocks are attributed “Granulate facies metamorphism,” and selected sedimentary rocks are attributed “Continental deposits”
- UNIT_UNCERTAINTY – a query following the map unit code indicates uncertainty about composition, or whether the rock is in situ
- MIN_AGE – minimum geologic age for the unit. Subdivisions of time-stratigraphic units are lower, middle, and upper (lower-case), and for plutonic rocks are Early, Middle, and Late

- MAX_AGE – see comments for MIN_AGE
- MIN_AGE_CODE – code derived from the geologic age codes defined by the AAPG Committee on Standard Stratigraphic Coding (1967)
- MAX_AGE_CODE – see comments for MIN_AGE_CODE
- AGE_UNCERTAINTY – a query preceding the map unit label indicates uncertainty about the assigned age
- MAP_UNIT_CODE – the GMNA map unit code
- MIN_MAX_RELATE – the relationship (“and”, “or”, “thru”) between the MIN?MAX ages of units bounded by multiple ages

REFERENCES

American Association of Petroleum Geologists, Committee on Standard Stratigraphic Coding (George V. Cohee, Chairman), 1967, Standard stratigraphic code adopted by AAPG: American Assiciation of Petroleum Geologists Bulletin, v. 51, no. 10, p. 2146-2150.

Reed, J.C. Jr., Wheeler, J.O., and Tucholke, B.E., compilers, 2004, Geologic Map of North America: Decade of North American Geology Continental Scale Map 001, Boulder, Geological Society of America, scale 1:5,000,000.

Reed, J.C. Jr., Wheeler, J.O., and Tucholke, B.E., 2005, Geologic Map of North America—Perspectives and explanation: Boulder, Colorado, Geological Society of America, Decade of North American Geology, 28 p.

Snyder, J.P., 1987, Map projections—a working manual: U.S. Geological Survey Professional Paper 1395, 383 p.

USGS National Surveys and Analysis Projects: Preliminary Compilation of Integrated Geological Datasets for the United States

By Suzanne W. Nicholson[1], Douglas B. Stoeser[2], Frederic H. Wilson[3], Connie L. Dicken[1], and Steve Ludington[4]

[1]U.S. Geological Survey
National Center, Mail Stop 954
Reston, VA 20192
Telephone: (703) 648-6344
Fax: (703) 648-6383
e-mail: swnich@usgs.gov
e-mail: cdicken@usgs.gov

[2]U.S. Geological Survey
Box 25046, Mail Stop 973
Denver, CO 80225
e-mail: dstoeser@usgs.gov

[3]U.S. Geological Survey
4200 University Dr.
Anchorage, AK 99508
e-mail: fwilson@usgs.gov

[4]U.S. Geological Survey
345 Middlefield Road, Mail Stop 901
Menlo Park, CA 94025
e-mail: slud@usgs.gov

INTRODUCTION

The growth in the use of Geographic Information Systems (GIS) has highlighted the need for regional and national digital geologic maps attributed with age and rock type information. Such spatial data can be conveniently used to generate derivative maps for purposes that include mineral-resource assessment, metallogenic studies, tectonic studies, human health and environmental research.

In 1997, the United States Geological Survey's Mineral Resources Program initiated an effort to develop national digital databases for use in mineral resource and environmental assessments. One primary activity of this effort was to compile a national digital geologic map database, utilizing state geologic maps, to support mineral resource studies in the range of 1:250,000- to 1:1,000,000-scale. Over the course of the past decade, state databases were prepared using a common standard for the database structure, fields, attributes, and data dictionaries. As of late 2006, standardized geological map databases for all conterminous (CONUS) states have been available on-line as USGS Open-File Reports. For Alaska and Hawaii, new state maps are being prepared, and the preliminary work for Alaska is being released as a series of 1:500,000-scale regional compilations. See below for a list of all published databases.

COMPILATION OF SPATIAL DATA FOR STATE GEOLOGIC MAPS

The first stage in developing state databases for the conterminous United States (CONUS) was to acquire digital versions of all existing state geologic maps. Although a significant number of digital state maps already existed, a number of states lacked them. For these states, new digital compilations were prepared by digitizing existing printed maps either in cooperation with the respective state geologic survey (e.g. OH, SD, TX) or by the USGS (e.g. KY, VT). In a few cases, we created digital state maps by merging existing larger scale digital files (e.g. SC, OK). It is important to note that, for this first round of compilation, we focused on compiling bedrock data for each state, although many state geological maps, especially in the West, combine both bedrock and surficial units on a single map.

All CONUS state databases were fit to a state boundary Arc/Info coverage, which was derived from the USGS 100k scale Digital Line Graphics (DLG) boundary layer quadrangles and has a polygon for each state. The purpose of fitting is so that adjoining state databases can be merged to form regional digital maps without slivers or overlaps at the state boundaries. Fitting was done by examining arcs along the boundary and extending or clipping them to the state boundary, depending on whether the arcs under or overshoot the boundary arc. No "rubber sheeting" was used. No attempt was made to reconcile differences in mapped geology between contiguous states. In the spatial tables, several fields were added in which a consistent set of terms was used for age and rock type, so that multiple spatial databases could be queried at the same time and allow generation of regional and national derivative maps based on age and rock type.

SUPPLEMENTAL ATTRIBUTE TABLES

The second stage was to assign values to a standard set of database fields in each state digital map database. Typically, state geologic maps contain more data than just arcs and polygons. Unit descriptions as well as age, lithologic, and bibliographic information present on the original source map were captured in a series of additional tables, including stratigraphic units, age, lithology, and references. For some older state map databases, more recent information was also captured. Figure 1 provides a schematic illustration of the structure of the spatial database tables and the supplemental tables. Figure 2 provides a more detailed look at the data entry format for the supplemental tables.

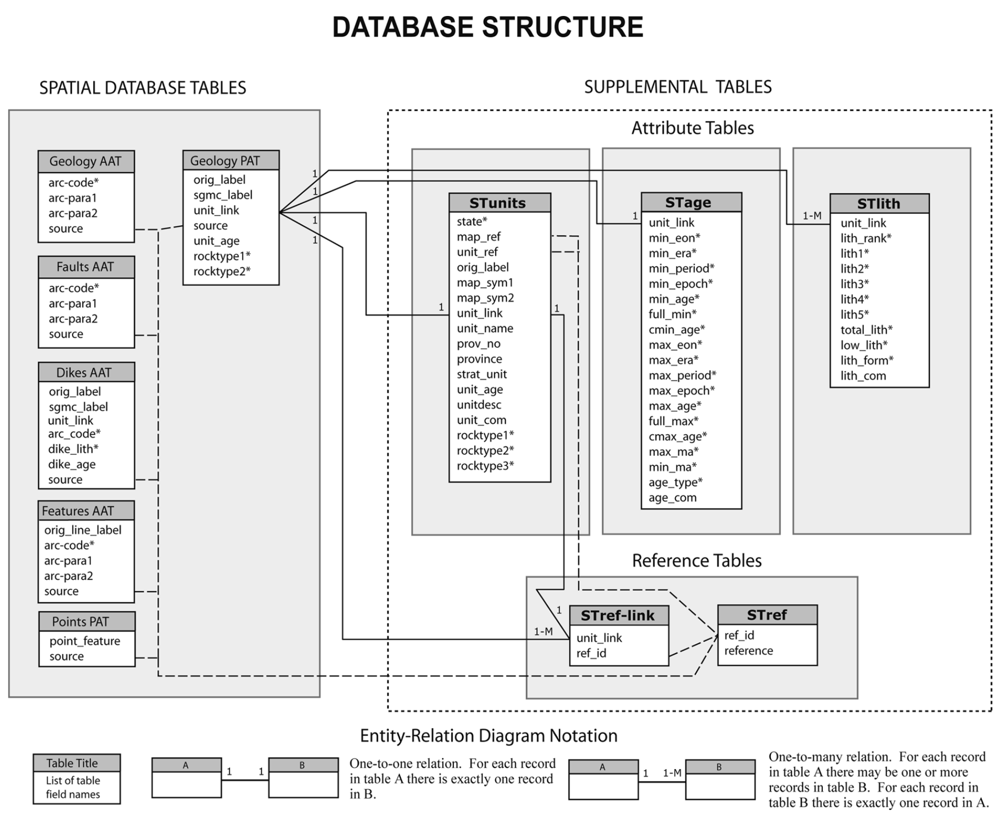


Figure 1. Data model for conterminous U.S. databases. Solid lines show links between tables through the unit_link field. Dashed lines indicate that the values in these fields have a one-to-one relationship to the ref_id field of the STref table. All values in the source, map_ref, and unit_ref fields are generated from the ref_id field in the STref table, though the definitions for each field are different. Fields marked with an asterisk are populated from a data dictionary. Figure is more legible in the web version.

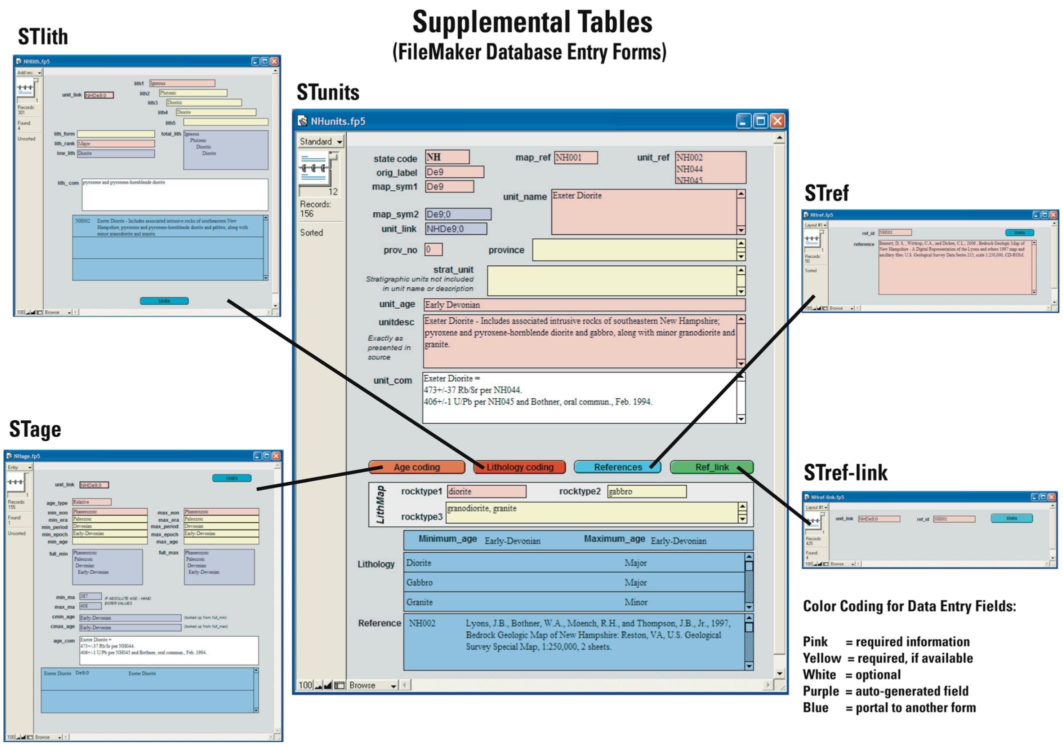


Figure 2. Illustration of the five supplemental tables (STunits, STlith, STage, STref, and STref-link, where ST stands for the two-letter abbreviation for each state). FileMaker 5, 5.5, or 6.0 was used to compile the supplemental tables, but it is not necessary to use this program in order to use the tables. Figure is more legible and in color in the web version.

STANDARD FILE SET

Conus

The files supplied for each state consist of (1) one or more spatial databases (Figure 1), and (2) a set of related supplemental tables (Figure 2). Each state database has the same database structure and attribution fields, which use terminology from standardized data dictionaries. At a minimum, the standard file set consists of a geology (polygon and arc, i.e. network coverage) spatial database, metadata, and supplemental attribute tables; however, additional spatial databases for other line or point features present on the source map may also be included (e.g., faults (when presented on the published maps), dikes, fold axes, volcanic vents, etc.). Detailed documentation of the standards, procedures, data dictionaries, and formats used accompanies each report.

The spatial databases are provided in ESRI export (.e00) and shapefile (.shp) formats. All spatial databases are provided both in geographic coordinates and a Lambert Conformal Conic projection for CONUS and geographic coordinates and UTM projection in Alaska, using a datum of NAD 27. The spatial database metadata are provided in three formats: ASCII text (.txt), Microsoft Word (.doc), and HTML (.htm). The supplemental data consist of related attribute tables (Figure 2): units (UNITS), age (AGE), lithology (LITH), and bibliographic references (REF). An additional table (REF-LINK) links spatial data and attributes to bibliographic references. The tables provide standardized attribution for the geologic map units for each map. These tables are available in comma-separated value (.csv), dBASE (.dbf), and FileMaker Pro (.fp5) formats, and for Alaska datasets as a runtime Filemaker Pro application. [Note, the .dbf format truncates all text fields at 256 characters, which impacts unit descriptions and, potentially, reference citations.]

Alaska and Hawaii

For Alaska and Hawaii, new state map compilations are being prepared. The data structure for Alaska is very similar, but not identical, to the data structure for the lower 48 states. The preliminary data for portions of Alaska are being released in a series of nominal 1:500,000-scale regional compilations. To date, ten new geologic compilations that cover more than two-thirds of the state of Alaska have been published (Figure 3). Additional compilations are currently being prepared for the

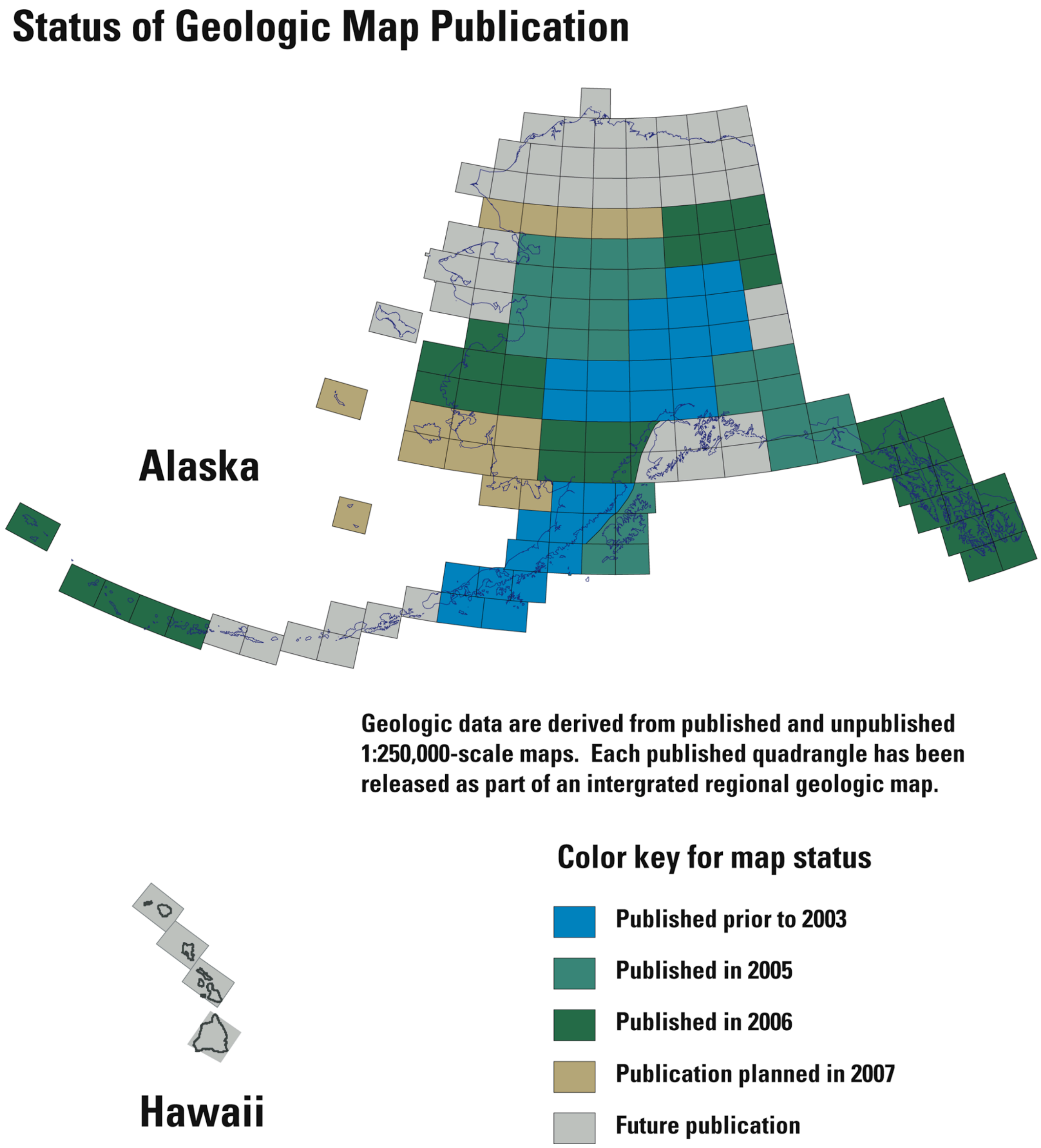


Figure 3. Map showing status of geologic map publication for quadrangles in Alaska and Hawaii. Figure is more legible and in color in the web version.

remaining portions of Alaska and for the state of Hawaii. Detailed documentation of the standards, procedures, data dictionaries, and formats used accompanies each report.

DERIVATIVE PRODUCTS OF THE STATE GEOLOGIC MAP COMPILATION

When the spatial databases are merged, these standardized tables allow development of derivative maps based on stratigraphy, lithology, and age. Figure 4 shows a map of the dominant rock type for each polygon, which was generated by plotting the controlled vocabulary values that appear in the *rocktype1* field. In Figure 5, the distribution of two rock types was generated by querying the *rocktype1* field in the state databases for *shale* and *granite*. A generalized geologic age map (Figure 6) was produced by generalizing the values in the free-form field *unit_age* in the spatial databases.

Another example of a use of these digital geologic state map data is the preliminary mineral resource assessment of North America, which is now underway by the USGS. The state geologic map datasets were used as base layers along with mineral occurrence data to outline tracts of favorable conditions for specific mineral deposit types.

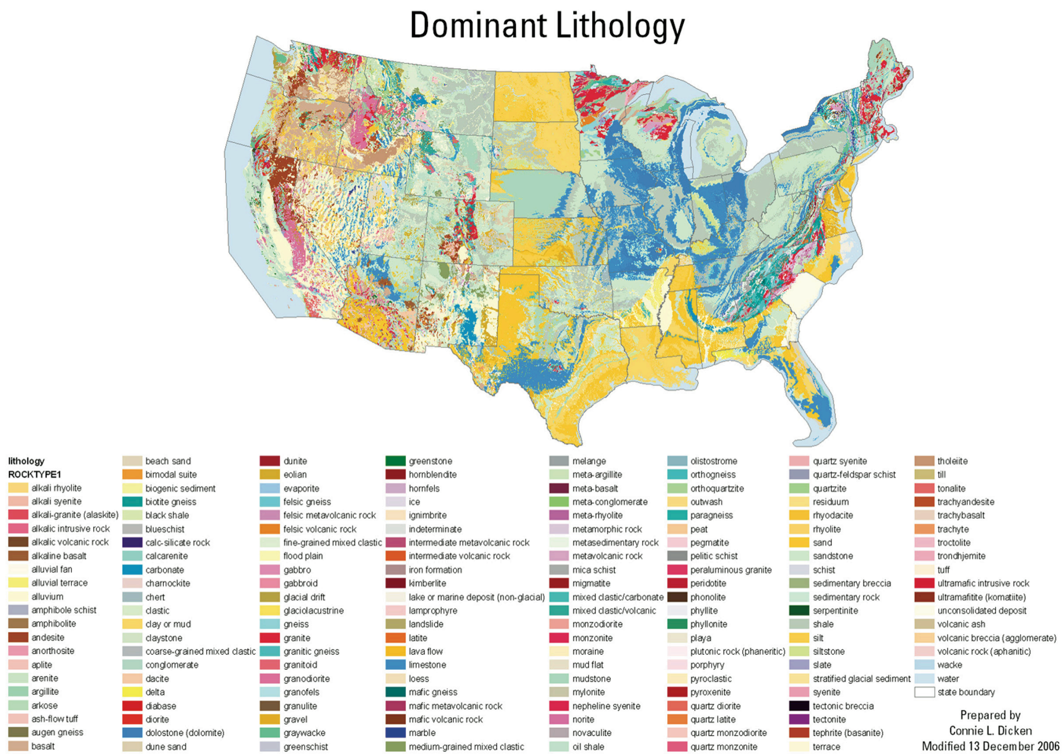


Figure 4. Distribution of dominant rock type for each polygon. Map was generated by querying the single-valued field *rocktype1*, which uses a controlled vocabulary. Figure is more legible and in color in the web version.

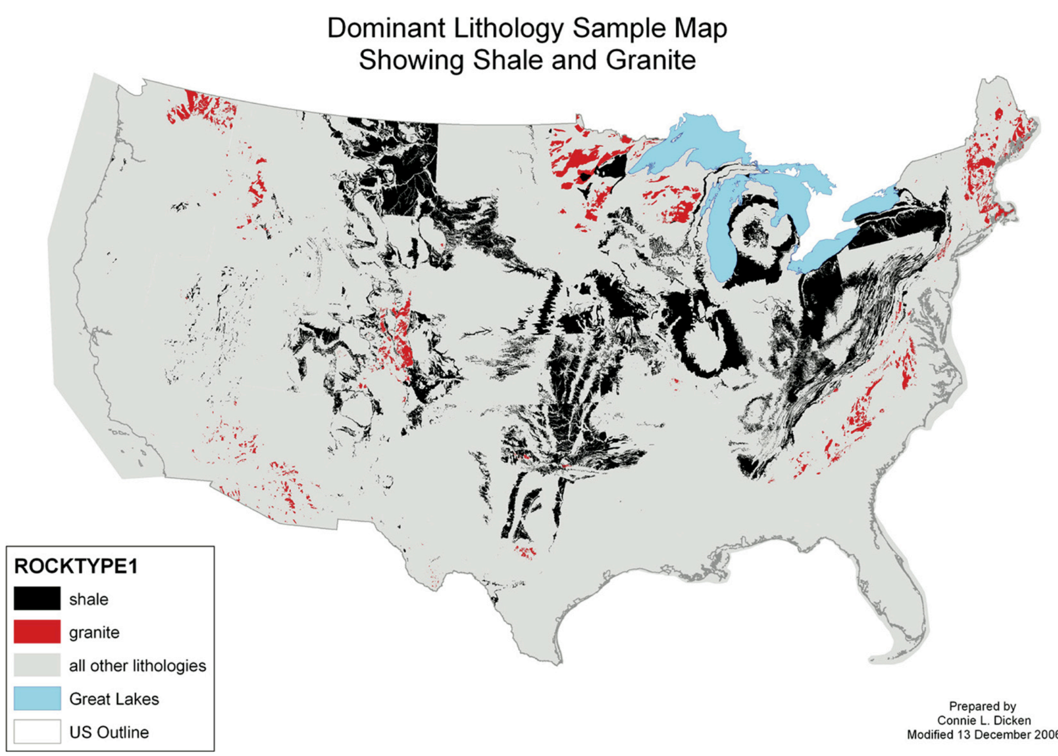


Figure 5. Map of the distribution of two rock types—shale and granite—was generated by querying the *rocktype1* field in the attribute table of the state databases for just those two rock types. Figure is more legible and in color in the web version.

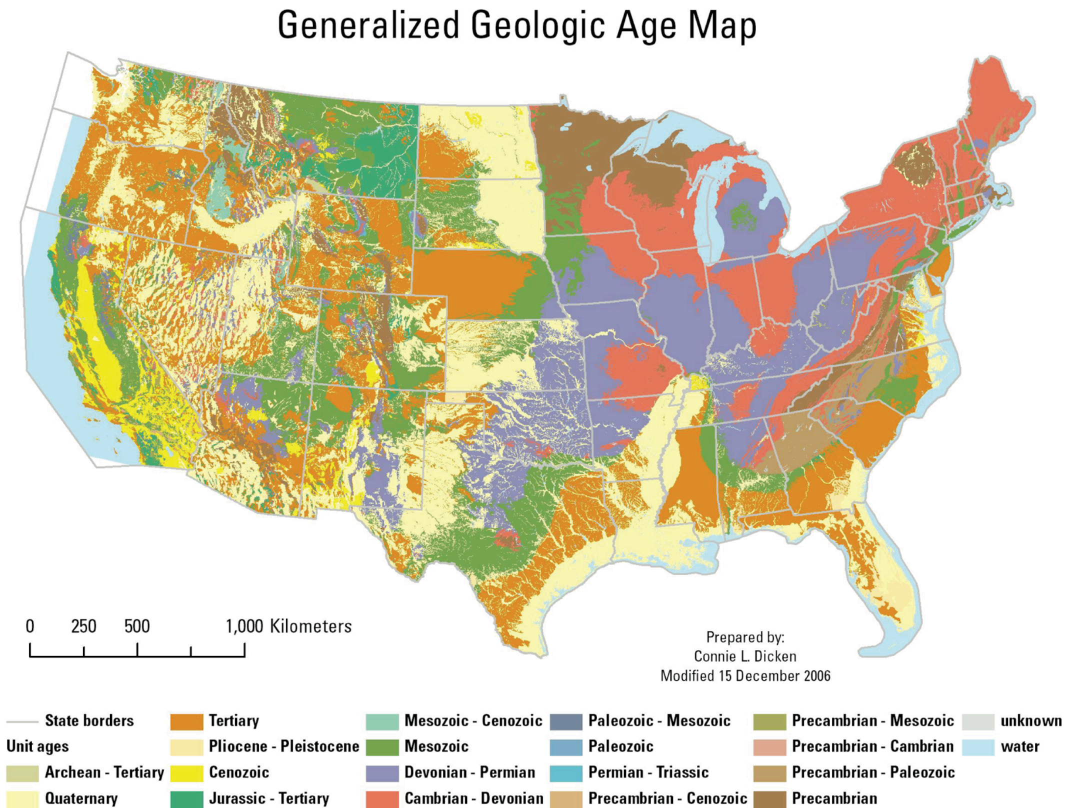


Figure 6. Generalized geologic age map produced by generalizing the values in the free-form *unit_age* field. Figure is more legible and in color in the web version.

BIBLIOGRAPHY OF DIGITAL STATE GEOLOGIC MAP COMPILATIONS

As of December 2006

Conterminous States:

Dicken, C.L., Nicholson, S.W., Horton, J.D., Foose, M.P., and Mueller, J.A.L., 2005, Preliminary Integrated Geologic Map Databases for the United States: Alabama, Florida, Georgia, Mississippi, North Carolina, and South Carolina: U.S. Geological Survey Open-File Report 2005-1323, on-line only, http://pubs.usgs.gov/of/2005/1323/.

Dicken, C.L., Nicholson, S.W., Horton, J.D., Kinney, S.A., Gunther, Gregory, Foose, M.P., and Mueller, J.A.L., 2005, Preliminary Integrated Geologic Map Databases for the United States: Delaware, Maryland, New York, Pennsylvania, and Virginia: U.S. Geological Survey Open-File Report 2005-1325, on-line only, http://pubs.usgs.gov/of/2005/1325/.

Ludington, Steve, Moring, B.C., Miller, R.J., Stone, P.A, Bookstrom, A.A., Bedford, D.R., Evans, J.G., Haxel, G.A., Nutt, C.J., Flyn, K.S., and Hopkins, M.J., 2005, Preliminary Integrated Geologic Map Databases for the United States: Western States: California, Nevada, Arizona, Washington, Oregon, Idaho and Utah, Version 1.2 (updated 12/01/06): U.S. Geological Survey Open-File Report 2005-1305, on-line only, http://pubs.usgs.gov/of/2005/1305/.

Nicholson, S.W., Dicken, C.L., Foose, M.P., and Mueller, Julie, 2004, Integrated Geologic Map Databases for the United States: The Upper Midwest States: Minnesota, Wisconsin, Michigan, Illinois, and Indiana: U.S. Geological Survey Open-File Report 2004-1355 on-line only, http://pubs.usgs.gov/of/2004/1355/.

Nicholson, S.W., Dicken, C.L., Horton, J.D., Labay, K.A., Foose, M.P. and Mueller, J.A.L., 2005, Preliminary Integrated Geologic Map Databases for the United States: Kentucky, Ohio, Tennessee, and West Virginia: U.S. Geological Survey Open-File Report 2005-1324, on-line only, http://pubs.usgs.gov/of/2005/1324/.

Nicholson, S.W., Dicken, C.L., Horton, J.D., Foose, M.P., Mueller, J.A.L., and Hon, Rudi, 2006, Preliminary Integrated Geologic Map Databases for the United States: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, Rhode Island and Vermont: U.S. Geological Survey Open-File Report 2006-1272, on-line only, http://pubs.usgs.gov/of/2006/1272/.

Stoeser, D.B., Green, G.N., Morath, L.C., Heran, W.D., Wilson, A.B., Moore, D.W., and Van Gosen, B.S., 2005, Preliminary Integrated Geologic Map Databases for the United States: Central States: Montana, Wyoming, Colorado, New Mexico, North Dakota, South Dakota, Nebraska, Kansas, Oklahoma, Texas, Iowa, Missouri, Arkansas, and Louisiana: U.S. Geological Survey Open-File Report 2005-1351, version 1.1 (updated 9/30/06) on-line only, http://pubs.usgs.gov/of/2005/1351/.

Alaska:

Gehrels, G.E. and Berg, H.C., 2006, Preliminary Integrated Geologic Map Databases for the United States: Digital Data for the Geology of Southeast Alaska: U.S. Geological Survey Open-File Report 2006-1290, on-line only, http://pubs.usgs.gov/of/2006/1290/.

Labay, K.A., Crews, Jesse, Wilson, F.H., Shew, Nora, and Hults, C.K., 2006, Preliminary Integrated Geologic Map Databases for the United States: Digital Data for the Generalized Bedrock Geologic Map, Yukon Flats region, East-central Alaska : U.S. Geological Survey Open-File Report 2006-1304, on-line only, http://pubs.usgs.gov/of/2006/1304/.

Richter, D.H., Preller, C.C., Labay, K.A., and Shew, N.B., compilers, 2006, Geologic map of the Wrangell-Saint Elias National Park and Preserve: U.S. Geological Survey Scientific Investigations Map 2877, scale 1:350,000, available online at http://pubs.usgs.gov/sim/2006/2877.

Wilson, F.H., Detterman, R.L., and Dubois, Gregory, 1999, Digital Data for the Geologic Framework of the Alaska Peninsula, Southwest Alaska, and the Alaska Peninsula Terrane: U.S. Geological Survey Open-File Report 99-317 on-line only, http://geopubs.wr.usgs.gov/open-file/of99-317/.

Wilson, F.H., Dover, J.H., Bradley, D.C., Weber, F.R., Bundtzen, T.K., and Haeussler, P.J., 1998, Geologic Map of Central (Interior) Alaska: U.S. Geological Survey Open-File Report OF 98-133-A, Version 1.2, available as paper copy, on-line, and CD, http://pubs.usgs.gov/of/1998/of98-133-a/.

Wilson, F.H., Labay, K.A., Mohadjer, Solmaz, and Shew, Nora, 2005, Preliminary Integrated Geologic Map Databases for the United States: Digital Data for the Reconnaissance Geologic Map of the Kodiak Islands, Alaska: U.S. Geological Survey Open-File Report 2005-1340 on-line only, http://pubs.usgs.gov/of/2005/1340/.

Wilson, F.H., Labay, K.A., Shew, Nora, and Hults, C.K., 2006, Preliminary Integrated Geologic Map Databases for the United States: Digital Data for the Reconnaissance Geologic Map of the lower Yukon River region, Alaska: U.S. Geological Survey Open-File Report 2006-1292, on-line only, http://pubs.usgs.gov/of/2006/1292/.

Wilson, F.H., Labay, K.A., Shew, Nora, Mohadjer, Solmaz, and Patton, W.W., Jr., 2005, Preliminary Integrated Geologic Map Databases for the United States: Digital Data for the Reconnaissance Geologic Map of the Yukon-Koyukuk Basin, Alaska: U.S. Geological Survey Open-File Report 2005-1341 on-line only, http://pubs.usgs.gov/of/2005/1341/.

Wilson, F.H., Labay, K.A., Shew, Nora, Preller, C.C., Mohadjer, Solmaz, and Richter, D.H., 2005, Preliminary Integrated Geologic Map Databases for the United States: Digital Data for the Geology of Wrangell-Saint Elias National Park and Preserve, Alaska: U.S. Geological Survey Open-File Report 2005-1342 on-line only, http://pubs.usgs.gov/of/2005/1342/.

Wilson, F.H., Mohadjer, Solmaz, Labay, K.A., and Shew, Nora, 2006a, Preliminary Integrated Geologic Map Databases for the United States: Digital Data for the Reconnaissance Geologic Map of the Western Aleutian Islands, Alaska: U.S. Geological Survey Open-File Report 2006-1302, on-line only, http://pubs.usgs.gov/of/2006/1302/.

Wilson, F.H., Mohadjer, Solmaz, Labay, K.A., and Shew, Nora, 2006b, Preliminary Integrated Geologic Map Databases for the United States: Digital Data for the Reconnaissance Bedrock Geologic Map for the Northern Alaska Peninsula area, Southwest Alaska: U.S. Geological Survey Open-File Report 2006-1303, on-line only, http://pubs.usgs.gov/of/2006/1303/.

Banding Birds with MapServer CGI

By Robert S. Wardwell and Kevin W. Laurent

U.S. Geological Survey
Cascades Volcano Observatory
1300 SE. Cardinal Court, Bldg. 10, Suite 100
Vancouver, WA 98683
Telephone: (360) 993-8908
Fax: (360) 993-8980
e-mail: rwardwell@usgs.gov

U.S. Geological Survey
Patuxent Wildlife Research Center
12100 Beech Forest Rd. Suite 4034
Laurel, MD 20708 - 4034
Telephone: (301) 497-5652
Fax: (301) 497-5826
e-mail: klaurent@usgs.gov

Since its establishment in 1936 as the nation's first wildlife experiment station, the USGS Patuxent Wildlife Research Center (PWRC) in Patuxent, Maryland, has been a leading international research institute for wildlife and applied environmental research. A primary goal at the PWRC Bird Banding Laboratory (BBL) is to manage the administration of bird banding permits, coordinate banding efforts, and collect data scientists can use to analyze such things as species behavior, migratory patterns, and the overall health of a species. Recently, computer scientists at the BBL have developed an ORACLE based permitting and banding administration system focused on tracking a wide range of data about banded birds and managing the BBL bird banding efforts (Figure 1). One component of this system relies on MapServer, an open source Web mapping solution, to collect an absolutely crucial set of data: the geographic location where a banded bird is encountered or recovered in the field. This paper describes how the BBL uses MapServer to collect scientific data and provides some insight into implementing MapServer.

MapServer was originally developed by the University of Minnesota ForNet project in cooperation with NASA and the Minnesota Department of Natural Resources. Presently, the MapServer project is hosted by the TerraSIP project, a NASA sponsored project between the UMN and consortium of land management interests. MapServer has become a popular tool by which to render spatial data such as vectors, maps, and images to the Web. The MapServer Website http://mapserver.gis.umn.edu/ is a valuable resource for learning more about MapServer.

MapServer's role in the BBL permitting and banding administration system is to generate interactive maps used to plot and record a geographic location where a bird was encountered or recovered. These data are currently captured at USGS call centers in Patuxent, Maryland, and Walla Walla, Washington, where reports are made when a banded bird is encountered or recovered in the field. USGS call center employees ask the caller to describe the bird's location and then use an interactive map to find the location and record the approximate coordinates. First, the caller describes a location using a common feature name, such as the name of a town, landmark, park, or water feature (Figure 2). Next, the BBL call center queries the system to search for the location in a feature names gazetteer that contains millions of common feature names in the US and abroad. Finally, MapServer is called upon to render an interactive map of the selected area, with tools that allow the call center employee to record the location where the bird was encountered (Figure 3). Since every caller is not equipped with a GPS or map, MapServer makes it possible to utilize the caller's relative proximity to known geographic features and calculate an accurate longitude and latitude values for their location.

Prior to using MapServer, these geographic data were collected by the BBL as a pair of vectors showing a spatial relationship between a known geographic feature and the location where a bird was encountered or recovered. For example, a banding encounter may be described as being "2 miles east and 3 miles north of Camas, Washington." These descriptions could be intersected and plotted within a mesh of 10 minute grid cells to add more quantitative measure to the data. This approach, however, led to

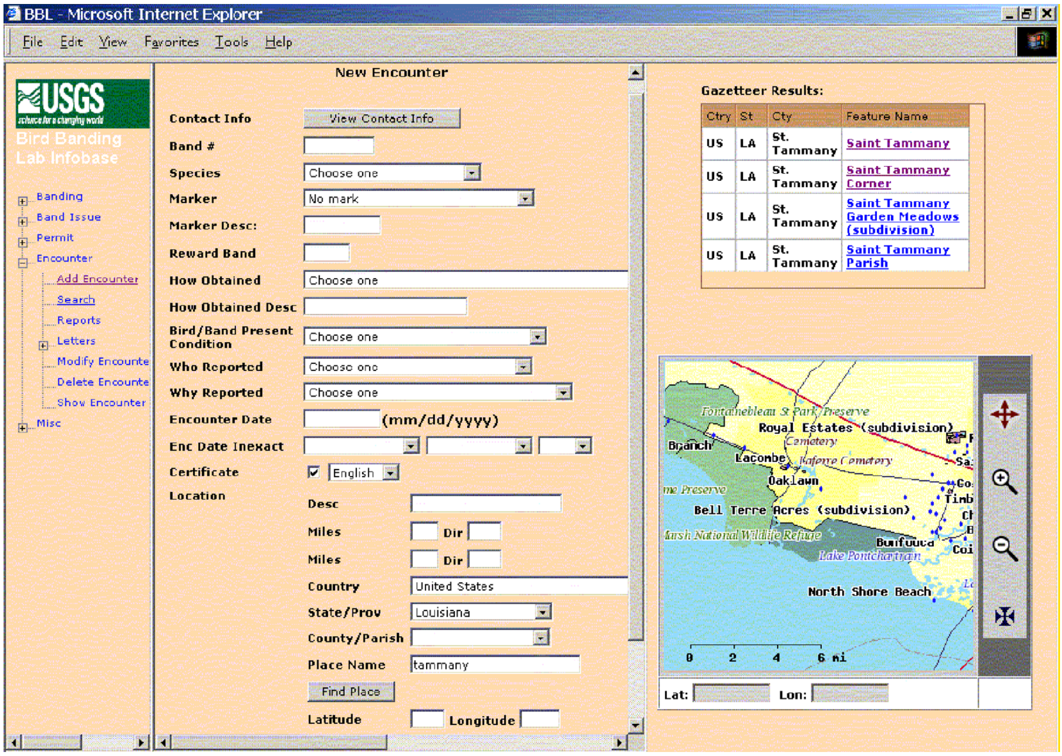


Figure 1. Screenshot of the banding management interface displays a data input form where information is collected about birds that are encountered in the field. A geographic search is initiated in the middle frame, which opens the MapServer interface in the right hand frame of the Web browser. The user identifies the location of the bird encountered by plotting it on the interactive map. Latitude and longitude coordinates are passed back to the main data entry form in the center of the page. Submitting the Web form records the event.

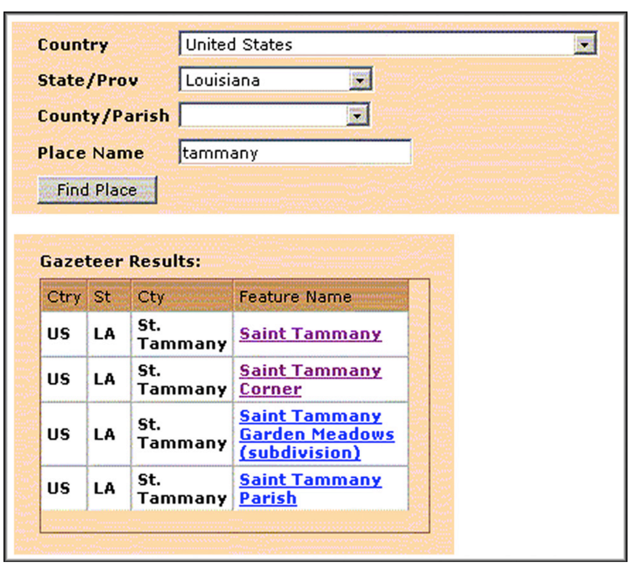


Figure 2. A place name search generates a list of possible name matches being returned in the *Gazetteer Results* list. Selecting a place name from the list passes the feature name coordinates to MapServer, which returns a map centered on the selected location.

a large degree of inaccuracy. Now, with the aid of feature rich maps rendered by MapServer, geographic location may be communicated more effectively and more accurate data is stored as coordinate pairs in the BBL database.

The maps generated by MapServer contain a selection of global and national spatial data layers. These layers include common landmarks, hydrology, urban areas, transportation routes, wildlife refuges, and parks for the United States along with Global political boundaries and populated places for the entire globe. Also included are transportation routes, hydrology, wildlife refuges, and parks for Canada and Mexico. MapServer is capable of serving both raster and vector data in a multitude of formats. Vector data sources, for example, may include ESRI shapfiles, PostGIS, ESRI ArcSDE, Oracle Spatial, MySQL. For the BBL banding application, base map data were downloaded from the Web in ESRI shapefile format. Feature data from the USGS Geographic Names Information System (GNIS), Natural Resources of Canada Gazetteer, and the National Geospatial Agency's Geographic Names System (NGA GNS) were made into shape files, which made it possible to search for and label over 3 million features on the maps. Quadtree based spatial indexing

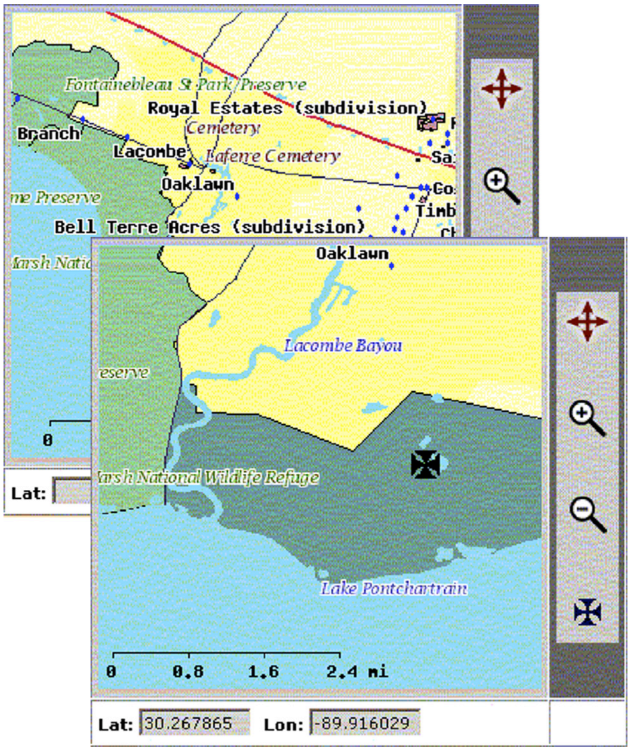


Figure 3. The interactive map created by MapServer contains tools used to zoom in, zoom out, pan, and mark a point on the map. In the above example, the user has zoomed in and clicked on the map, which places an "x marks the spot" symbol at the location where a bird was encountered. MapServer calculates the coordinates of the point which is then submitted to the database.

```
LAYER
  NAME "USParks"
  DATA parks_shp
  STATUS default
  TYPE POLYGON
  MAXSCALE 2000000
  LABELMAXSCALE 2000000
  LABELITEM "Name"
  CLASS
    COLOR 153 204 153
    OUTLINECOLOR 0 0 0
    LABEL
      SIZE 8
      TYPE TRUETYPE
      POSITION auto
      FORCE false
      COLOR 51 102 0
      OUTLINECOLOR 255 255 255
      FONT lucida-bright-italic
    END
  END
END
```

Figure 4. The *map* file (.map extension) is a text file created manually or built using one of many open source tools. This file describes the data sources for your map and defines the map extent, data layers, symbols, and layer classification.

of each shape file was performed using the MapServer `shptree` utility, which helps speed the delivery of map data over the Web.

In the BBL banding application, MapServer operates as a CGI program that handles requests and responses. The map and everything about the layers within the map, such as cartographic symbols and the classification of the layers, are controlled through a MapServer *map* file. This file is a hierarchical text file with a .map extension that describes each of the data layers to be included in the map, and describes how each layer is to appear (Figure 4). The interactive mapping interface, which contains tools used to zoom, pan, or plot the location of a banded bird encounter, is controlled by an HTML and JavaScript template file. The template file contains unique CGI variable tags, which are handled by MapServer each time the user makes a request for a new map (Figure 5). The map itself is rendered onto the Web page as a static image, such as a .gif, .png, or .jpg that is replaced when MapServer processes a new set of input parameters from the HTML template.

When the BBL permitting and band management application went to production use in March 2006, the mapping component contained over 50 data layers, which included data for 3.6 million named points in more than 50 countries. MapServer proved itself worthy by handling these large datasets and displaying them efficiently over the Internet. Furthermore, we found MapServer to be a simple, elegant solution for improving the data acquisition at the BBL. The success of this project can be attributed to the collaborative efforts of biologists, computer scientists, and GIS professionals, along with the open source development community surrounding MapServer.

REFERENCES

Kropla, B., 2005, Beginning MapServer-Open Source GIS Development: Apress Publishing.

Lime, S., (from the MapServer Website), 2006, Welcome to MapServer, accessed at http://mapserver.gis.umn.edu/.

Mitchell, T., 2005, Web Mapping Illustrated - Using Open Source GIS Toolkits: O'Reilly Publishing.

Patuxent Wildlife Research Center Website, 2006, Patuxent Wildlife Research Center – Mission: accessed at http://www.pwrc.usgs.gov/aboutus/mission.cfm.

```
<h3>USGS Breeding Bird Survey Interactive Route Map</h3>
<form action="/cgi-bin/mapserv" name="mapForm" method="get" target="_self" onsubmit="return(map_Submit(this))">
   <input type="hidden" name="map" value="[map]">
   <input type="hidden" name="imgext" value="[mapext]">
   <input type="hidden" name="mode" value="browse">
   <input type="hidden" name="zoomdir" value="[zoomdir]">
   <input type="hidden" name="zoomsize" value="[zoomsize]">
   <input type="hidden" name="layers" value="[layers]">
   <input type="hidden" name="imgxy" value="149.5 149.5">
   <input type="hidden" name="imgbox" value="-1 -1 -1 -1">
<table width="460" border="1" cellspacing="0" cellpadding="4" align="left" bgcolor="#666666">
 <tr>
  <td align="center" bgcolor="#CCCCCC" valign="top">
   <input property="" type="image" name="img" src="[img]" width="400" height="550" border="0" alt="map image">
  </td>
 <td>
```

Figure 5. The HTML template file defines a map interface that can be customized and enhanced with JavaScript. The MapServer CGI program will process your map file and pass values to “substitution strings” that are enclosed in square brackets (‘[ ]’). The CGI variable examples in this code snippet include `map`, `mapext`, `zoomdir`, `zoomsize`, `layers`, and `img`.

APPENDIX A

List of Workshop Attendees

[Grouped by affiliation]

Alaska Division of Geological and Geophysical Surveys
Kenneth Papp

Arizona Geological Survey/USGS
Stephen Richard

Arkansas Geological Commission
Jerry Clark
Doug Hanson

Avenza Software
David Andrec
Doug Smith

British Geological Survey
Jeremy Giles

Colorado Geological Survey
Beth Widmann

Colorado State University - National Park Service
James Chappell
Ron Karpilo
Stephanie O'Meara
Trista Thornberry-Ehrlich

Czech Geological Survey
Zuzana Krejci
Robert Tomas

ESRI
Jaynya Richards

Geological Survey of Alabama
Philip Patterson

Geological Survey of Canada
Eric Boisvert
Boyan Brodaric
Peter Davenport
David Viljoen

Hydrogeologic Services
Curtis Coe

Idaho Geological Survey
Jane Freed
Loudon Stanford

Illinois State Geological Survey
Sheena Beaverson
Jennifer Carrell
Jane Domier
Deette Lund

Indiana Geological Survey
Rick Hill
Kim Sowder
Renee Stubenrauch

Iowa Geological Survey
Jim Giglierano

Kansas Geological Survey
John Dunham
William Harrison

Kentucky Geological Survey
Jerry Weisenfluh

Louisiana Geological Survey
R. Hampton Peele

Manitoba Geological Survey
Greg Keller

Michigan Geological Survey
John Esch

Minnesota Geological Survey
Harvey Thorleifson

Missouri Geological Survey
Edith Starbuck

National Park Service
Tim Connors

Natural Resources Canada
Vic Dohar
Dave Everett
Gary Grant
Linda Guay
Mario Methot

Nevada Bureau of Mines and Geology
Jennifer Mauldin

New Hampshire Geological Survey
Rick Chormann

New Mexico Bureau Geology & Mineral Resources
Glen Jones

North Carolina Geological Survey
Jeff Reid

Oglebay Norton
Steve Murdoch

Ohio Department Of Natural Resources-Division of Water
Mike Angle
Tim Beck
Wayne Jones

Ohio Environmental Protection Agency
Christopher Kenah
David White

Ohio Geological Survey
Tom Berg
Bruce Gerke
Don Guy
Dennis Hull
Ed Kuehnle
Connie Livchak
Jim McDonald
Rick Pavey
Donovan Powers
Doug Shrake
Mac Swinford
Erik Venteris
Joseph Wells

Ohio State University
Kelly Barrett
Michael Bevis
Kelly Carroll
Kathryn Epp
Mike Fidler
Steve Goldsmith
Tim Hawthorne
Garry McKenzie
Xutong Niu
Karen Royce
Alan Saalfeld
Frank Schwartz
Mary Scott

Ontario Geological Survey
Zoran Madon

Optimal Geomatics, Inc.
Mark Brooks

Oregon Dept. of Geology & Mineral Industries
Paul Staub

Pennsylvania Geological Survey
Jay Parrish
Tom Whitfield

Portland State University
Dave Percy

Reserve Energy Exploration, Inc.
William Haas

South Carolina Geological Survey
Erin Hudson

State of Ohio/OIT/SDD/ESS
Stuart Davis

The Nature Conservancy
August Froehlich

U.S. Geological Survey
Stafford Binder
Luke Blair
Peter Chirico
Connie Dicken
Joseph East
Chris Garrity
Charles Hickman
Bruce Johnson
Victor Mossotti
Suzanne Nicholson
Dave Soller
Nancy Stamm
Rob Wardwell
Ric Wilson

University of Alabama
Doug Behm

University of Arizona/USGS
Jonathan Craigue

Harry McGregor

USDA Forest Service
Andy Rorick

Utah Geological Survey
Kent Brown

West Virginia Geological Survey
Gayle McColloch

Jane McColloch

Wisconsin Geological & Natural History Survey
Michael Czechanski
Deborah Patterson
Peter Schoephoester

Wyoming State Geological Survey
Allory Deiss
David Lucke
Phyllis Ranz