From "Digital Mapping Techniques '08—Workshop Proceedings"
U.S. Geological Survey Open-File Report 2009–1298
http://pubs.usgs.gov/of/2009/1298/

Geologic Content Specification for a Single-Map Database

By Stephen M. Richard[1] and David R. Soller[2]

[1] Arizona Geological Survey / U.S. Geological Survey
416 W. Congress #100
Tucson, AZ 85701
Telephone: (520) 770-3500
Fax: (520) 770-3505
email: *steve.richard@azgs.az.gov*

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

Introduction

The design of geoscientific database systems for use by geological surveys presents a special set of challenges. In contrast to application-specific designs that are driven by a particular use case, geological survey systems need to allow for the wide variety of data typically collected, and for the possibility of new projects demanding new types of content. The published and unpublished information archive function of most geological surveys establishes a need to record the source and update history of that information. Because users have widely varying needs and technical capabilities, a simple and easily understood data structure is most practical, because the budgets of most surveys do not provide a level of funding necessary to implement and support sophisticated, or multiple, data management systems. The geological survey community has long recognized the potential benefits of a standardized database platform to enable savings by sharing tools, training, and documentation costs. Several approaches to developing a community database schema for geoscience information have been defined and evaluated in the past 15 years. These include a relational database design (e.g., "v.4.3" of Johnson and others, 1998), a conceptual model ("NADM C1" of NADMSC, 2004), and numerous other database designs by various agencies involved in geologic map production, documented to varying degrees and related to each other in various ways. No consensus design has emerged.

The U.S. National Geologic Map Database project (NGMDB, *http://ngmdb.usgs.gov/*) has participated in this development and testing of database designs and has test-implemented a complex enterprise database design based on the NADM C1 model (Richard and others, 2004). In light of comments made above, we are continuing to work toward a relational database schema that achieves a functional balance between simple design and ability of the data structure to accurately represent some of the complex relationships inherent in geoscience information. Our objective is to develop a database that is primarily useful for delivering a single dataset (e.g., a publishable geologic map), thereby providing a moderate level of functionality that will be useful for data interchange.

Some Proposed Designs

We have surveyed a variety of logical designs that have been proposed or used over the years for geoscience databases (see Selected References). These range from a simple flat-file data structure to complex, highly normalized relational (database) or network (XML) schema. Based on the common geoscience entities in these schemes, we have developed a content specification that we propose as the basis for a geoscience delivery database.

Except for simple, single flat-file designs, all the designs involve a link from a spatial object (point, line, or polygon) to one or more descriptive entities that are collections of properties, and some scheme for associating the spatial objects with symbols for portrayal. There is some variation in the logical structure for associating the spatial objects with multiple descriptions, for associating descriptions with classes or instances, and for the amount of location, observation method, and confidence metadata associated with the spatial object location and classification.

Given that the various logical models are broadly compatible, the major challenge to developing a basis for information interoperability is to agree on the content model or specification for the kinds of information that should be included in a geoscience database. The NADM C1 conceptual model is a content model, but it uses Unified Modeling Language (UML) terminology that is unfamiliar to many geoscientists. The CGI Interoperability Working Group (Commission for the Management and Application of Geoscience Information (CGI), a commission of the International Union of Geological Sciences (IUGS)) built on the NADM C1 model to develop an XML markup schema for geoscience information (*https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/GeoSciML*). The GeoSciML XML schema is currently being tested and implemented by various geological surveys. This paper is a distillation of these various geoscience data models to outline a broadly applicable content specification for a geoscience database.

Content Specification

A content specification is a listing of the kinds of information that need to be represented in a table, an XML element, or an entire database. The content specification informs users on the information content of a resource and helps software designers to understand the scope and requirements of tools that will be used to maintain or access the resource.

In the simple content specification presented here, principal geoscientific and cartographic elements are described. It is basically a conceptual model expressed in natural language. More detailed and proscriptive specifications are necessary to allow development of software that would access the design. Those specifications would be based on this simple content model. The content specification we outline here has three potential applications:

1. Facilitating data interchange – A shared data schema for transferring packages of geoscience information, but not visualization, for input into automated processing pipelines that use geoscience data, or for publishing data in an implementation-independent format. Because the design anticipates automated processing, it requires a very proscriptive specification, e.g. XML schema and controlled vocabularies from a registry.
2. Off-the-shelf package for geoscience GIS – A geodatabase against which standard tools could be developed for automating geologic map data (e.g., ArcGIS extension), and to facilitate the use of maps as an aid for inputting observation data. This would be most useful for agencies that are not already using a geodatabase-based approach, or are new to GIS. This application would need topology rules to assist line and polygon editing, and domains to usefully restrict user input to minimize errors in data entry. For the long term, tools to simplify construction of the map collar legend and symbolization scheme would be feasible.
3. Map viewing– A mechanism for packaging a complete map layout (e.g., the cartography) along with the associated data. The spatial objects must have a simple linkage to symbols used to depict them. Mechanisms for implementing transport of actual graphic objects (e.g., ArcMap layer, ArcView legend file, OCG styled layer descriptor file) are out of scope but would be necessary for a working system.

Proposal for Geologic Map Dataset Content

The content listed here is a summary of the content related to a typical geologic map publication. Each agency or community would need to evaluate the importance of these different content elements to determine if they should be optional or required for information interchange.

Geologic Units

Geologic units are identifiable, mappable parts of the Earth (North American Geologic-Map Data Model Steering Committee, 2004). A geologic unit description may be associated with a unit distinguished by a particular symbol (color) on a geologic map (a map unit), or with a subset of polygons assigned to a map unit in order to describe variations in that unit, or with individual observation locations and associated raw field data. Suggested content is listed below. Table 1 contains a supplementary list of properties associated with geologic units in the GeoSciML model, as examples of other more detailed information that might be included.

- Stratigraphic Unit – association of a mapped unit with some formal nomenclature scheme.
- Lithologic composition – representation of the composition of a geologic unit in terms of the kinds of rocks that form the unit.
- Preferred age – assignment of a single geologic age, chosen to be most representative of the mapped unit. The age specification can be numeric or based on named eras from a stratigraphic time scale.

- History of unit (sequence of events) – a sequence of events in the formation of the mapped unit. The preferred age will be the age of one of these events.
- Geomorphic character – the Earth surface expression of the mapped unit.
- Symbolization – association of the mapped unit with a graphical element for the purpose of map display.
- Metadata – information documenting the provenance of the geologic unit description.
- Thickness – thickness of the mapped unit, may be reported as a single value or a range.

Marker Beds

Marker beds are a kind of geologic unit included in the North American Code of Stratigraphic Nomenclature as the lowest-level unit. Marker beds that are not unit boundaries do not participate in the topology, but are always contained within some other stratigraphic unit of which they are a part.

Geologic Structures

A geologic structure is a configuration of matter in the Earth based on describable inhomogeneity, pattern, or fracture (North American Geologic-Map Data Model Steering Committee, 2004). A geologic structure description may be associated with one to many individual features. Suggested content is listed below. Table 2 contains a supplementary list of properties associated with geologic structures in the GeoSciML model, as examples of other more detailed information that might be included.

- Type or category classifier – association of the mapped feature with some category that describes the geologic nature of the feature.
- Preferred age – assignment of a single geologic age, chosen to be most representative of the geologic structure. The age specification can be numeric or based on named eras from a stratigraphic time scale.
- History (sequence of events) – a sequence of events in the formation of the structure. The preferred age will be the age of one of these events. For faults, this property would include the slip history across the fault.
- Orientation – for a planar or linear structure, geologists are interested in the orientation of the feature in an earth-surface reference frame, typically relative to a horizontal surface with azimuth measured relative to north.
- Symbolization – association of a graphical element with the structure, for purposes of map display.
- Metadata – information documenting the process by which the structure was located, who mapped the feature, and less often, the context in which the feature was mapped.

Fold-Hinge Surface Traces

In terranes with map-scale folds, the traces of hinge surfaces commonly are mapped to help elucidate the geometry and location of the folds. These surfaces are based on the orientation of bedding or the curvature of traceable folded surfaces such as marker beds or contacts between units.

Table 1. Selected list of additional properties associated with geologic units, from the GeoSciML specification.

Bedding pattern	Magnetic susceptibility	Porosity
Bedding style	Metamorphic facies	Unit thickness
Bedding thickness	Metamorphic grade	Weathering degree
Body morphology	Outcrop character	Weathering process
Composition category	Peak pressure value	Weathering product
Contained structure	Peak temperature value	Weathering environment
Density	Permeability	
Exposure color	Protolith rock type	

Table 2. Selected list of additional properties associated with geologic structure descriptions, from the GeoSciML specification.

amplitude	hingeLineCurvature	profileType
axialSurfaceOrientation	hingeLineOrientation	segment
boundedGeologicUnit	hingeShape	spacing
contactCharacter	intensity	span
continuity	interLimbAngle	symmetry
definingElement	layerComposition	totalDisplacement
foldSystemMember	limbShape	wavelength
geneticModel	mineralElement	
higherOrderFoldPart	periodicity	

Dikes and Veins

Dikes and veins are sheet-like bodies of intruded rock, which are generally too thin to represent as polygons. Dikes and veins may have multiple disconnected instances (e.g., outcrops), unified by their lithologic character and relationship to other units. Dikes and veins are intruded along fractures, and thus have a structural aspect as well. Although they may intrude along contacts or faults, dikes and veins generally are independent of the topology of surfaces bounding the principal geologic units in an area.

Observation Points

Field data that are the basis for compiling geologic maps generally are acquired through observations collected at point locations, which then are extrapolated to define the map geometry. Point observations may be associated with map unit descriptions, structure descriptions, or samples. The observation points and original, detailed descriptive data typically are not included directly in a map interpretation, but are important to store in geoscience databases because they document the fundamental observational basis for the interpretations represented on a map.

Sample Locations

Physical samples collected in the field are referenced to their geologic setting by specifying the location at which they were collected, possibly along with more detailed observation data of relationships between the sampled material and the surrounding rock and structure at the sample location.

Cartographic Points

The geologic map includes numerous annotation items that could be located by points in the map layout. These include graphical symbols that encode arrows showing fault dip direction, bar and ball or other symbols that encode fault slip or separation sense, symbols used to indicate the kind of fold closure observed along a mapped hinge surface trace, numbers to indicate magnitude of dip or plunge of structures, map unit labels, or generalized, representative measurements of strike and dip for planar structures such as bedding. These are called cartographic points because their location may be determined by graphical and esthetic considerations, in order to best communicate some aspect of the geologic information in a particular map portrayal. In ESRI Geodatabase language, these would be included in one or more annotation feature classes, which include a bounding box geometry and encode the graphical element in an opaque 'blob' field. A more transparent (but less flexible and functional) approach is to include one or more point feature classes to locate structure symbols, label text for polygons, and label text for inclination values of structure data in a map layout.

Spatial Data and Database Feature Classes

Because geologic features are located in the Earth, a fundamental content requirement for any geoscience database is the ability to accurately represent locations in the Earth. Current geographic information system software is designed to represent and manipulate descriptions of location. Two-dimensional systems typically allow representation of points, lines (ordered collections of points), and polygons (collections of lines that form closed rings). It is anticipated that in the future these systems will operate in three dimensions as well, and the feature types will expand to include volumes.

Geoscience features are associated with spatial objects in order to place them in a geographic context. One of the basic design decisions that must be made in any geologic spatial database is the mapping between geoscience features and database feature classes. The term "feature class" is used here to mean a representation entity (e.g., a table in a database, an element in an XML document) that has a location property. Criteria affecting the feature class design include entity typing based on properties associated with the entity, and the relationships between entities that are built into the representation schema to enable various use cases for the database implementation (e.g., digitizing and editing, data archive, quality assurance, cartographic design). At a conceptual or logical level, it is common to try to preserve a correspondence between the feature classes and geoscientific entities. For physical implementations, pragmatic design criteria become of primary importance. These include ease of implementation and use, and constraints that may be determined by the specific GIS software environment.

In general, points in geoscience geographic databases locate observations or sample collections, or more rarely they represent geologic features that are too small to represent with an extended geometry (line, polygon) at the scale of representation. Lines locate the intersection of planar geologic features with the map horizon represented by the map portrayal. Thus, lines typically are associated with contacts between units, faults, dikes, veins, and marker beds. Lines also may represent geologic features that are inherently linear, such as scarp crests, dune crests, channel axes, or fold hinges in a particular folded surface. Polygons represent patches within which a mapped unit is found on the map horizon. Typically the mapped unit is a rock volume, and the polygon represents the intersection of a three-dimensional body with the map horizon. Less commonly, the polygon may represent a unit defined by the character of the outcrop surface. The specific partitioning of these representations into different feature classes is outside the scope of this specification.

Conclusions

Design of a widely useful community geoscience database schema depends on careful consideration of the purpose of the schema, and scoping of the content to balance simplicity against depth of scientific representation. The specification presented here is based on comparison of existing database implementations as a means of identifying content that is commonly included in actual datasets. This sort of ‘bottom-up’ approach has become increasingly useful as more geoscience databases have been implemented and populated, and provides an instructive counterpoint to the more ‘top-down’ design processes that have dominated proposals for standardized schemas.

Selected References

Bain, K.A. and Giles, J.R.A., 1997, A standard model for storage of geologic map data: Computers and Geosciences v. 23, p. 613-620.

Brodaric, B., Journeay, M., Talwar, S., and others, June 18, 1999, CordLink Digital Library Geologic Map Data Model Version 5.2: Geological Survey of Canada unpublished report, accessed at *http://cordlink.gsc.nrcan.gc.ca/cordlink1/info_pages/English/dm52.pdf*, accessed June 13, 2001.

Brodaric, Boyan, and Hastings, Jordan, 2002, An object model for geologic map information, *in* Richardson, D., and van Oosterom, P., eds., Advances in Spatial Data Handling, 10th International Symposium on Spatial Data Handling: Heidelberg, Germany, Springer-Verlag, p. 55-68.

Gautier, D.L., 1999, Data model for single geologic maps: An application of the national geologic map data model, *in* Soller, D.R., ed., Digital Mapping Techniques ’99—Workshop Proceedings: U.S. Geological Survey Open-File Report 99-386, p. 77-81, *http://pubs.usgs.gov/of/1999/of99-386/gautier.html*.

Harris, Carl F.T., 1998, Washington State’s 1:100,000–scale geologic map database: An ArcInfo data model example, *in* Soller, D.R., ed., Digital Mapping Techniques ‘98—Workshop Proceedings, U.S. Geological Survey Open–File Report 98–487, p. 27–35, *http://pubs.usgs.gov/openfile/of98–487/harris.html*.

Hastings, J.T., Raines, G.L., and Moyer, L.A., 2008, Proposal for an ArcGeology Version 1 – A geodatabase design for digital geologic maps using ArcGIS, *in* Soller, D.R., ed., Digital Mapping Techniques ‘07—Workshop Proceedings: U.S. Geological Survey Open-File Report 2008-1385, p. 59-72, *http://pubs.usgs.gov/of/2008/1385/pdf/hastings.pdf*.

Johnson, B.R., Brodaric, B., and Raines, G.L., 1998, Digital Geologic Maps Data Model, V. 4.3: unpublished report, accessed at *http://www.nadm-geo.org/dmdt/*.

Lewis, R.S., Burmester, R.F., Kauffman, J.D., and Frost, T.P., compilers, 2001, Digital geologic map of the St. Maries 30 x 60 minute quadrangle, Idaho: Idaho Geological Survey, Digital Geologic Map 1, version 2.2002.06.

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 04-1334, 61 p., *http://pubs.usgs.gov/of/2004/1334/*.

O’Meara, S.A., and Poole, A.R., 2004, National Park Service Geologic Resources Inventory Geology–GIS coverage shapefile data model: unpublished report, 107 p., accessed at *http://www1.nature.nps.gov/im/gis/GeologyGISDataModel.htm*.

O’Meara, S.A., Stanton, H.I., Chappell, Jim, and others, 2005-2007, NPS GRI Geology-GIS Geodatabase Data Model (V. 1.1 to 1.4): unpublished report, 147 p., accessed at *http://www1.nature.nps.gov/im/gis/GeologyGISDataModel.htm*.

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.

Richard, S.M., 2003, Geologic map database implementation in the ESRI Geodatabase environment, *in* Soller, David R., ed., Digital Mapping Techniques '03—Workshop Proceedings: U.S. Geological Survey Open-File Report 03-471, p. 169-183, *http://pubs.usgs.gov/of/2003/of03-471/richard2/index.html*.

Richard, S.M., and Orr, T.R., 2001, Data structure for the Arizona Geological Survey Geologic Information System – Basic geologic map data, *in* Soller, David R., ed., Digital Mapping Techniques '01—Workshop Proceedings: U.S. Geological Survey Open-File Report 01-223, p. 167-188, *http://pubs.usgs.gov/of/2001/of01-223/richard2.html*.

Richard, S.M., Craigue, J., and Soller, D.R., 2004, Implementing NADM C1 for the National Geologic Map Database, *in* Soller, D.R., ed., Digital Mapping Techniques '04—Workshop Proceedings: U.S. Geological Survey Open-File Report 2004-1451, p. 111-144, *http://pubs.usgs.gov/of/2004/1451/pdf/richard.pdf*.

Richard, S.M., and CGI Interoperability Working Group, 2007, GeoSciML – a GML application for geoscience information interchange, *in* Soller, D.R., ed., Digital Mapping Techniques '06—Workshop Proceedings, U.S. Geological Survey Open-File Report 2007-1285, p. 47-59, *http://pubs.usgs.gov/of/2007/1285/pdf/Richard.pdf*.

Soller, D.R., Brodaric, Boyan, Hastings, J.T., Wahl, Ron, and Weisenfluh, G.A., 2002, The central Kentucky prototype: An object-oriented geologic map data model for the National Geologic Map Database: U.S. Geological Survey Open-File Report 02-202, 39 p., *http://pubs.usgs.gov/of/2002/of02-202/*.