Prototype Implementations of the North American Data Model Steering Committee Data Model for a Geologic Map Database

by Ronald R. Wahl, U.S. Geological Survey, Denver, Colorado, USA

ABSTRACT

The National Geologic Map Database Project (NGMDB) of the U.S. Geological Survey (USGS) has been working to implement the North American Data Model Steering Committee (NADMSC) data model in a number of software systems. The NGMDB has implemented a number of prototype databases using object-relational software. Additional benefits discovered from this work are: iterative changes to the underlying data model do not change the data already in place, implementation of dynamic map generation based on map unit generalizations is straight-forward, and the geologic features are not necessary referred to by the geometry of the feature. Some of the problems encountered so far are: geologic language tables to accompany the database have been difficult to develop, a list of routine queries for use with the database have been slow to evolve, and the routine importation of multiple data formats from the same GIS software has not been solved.

Introduction

History

Since enacting of the Geologic Mapping Act (GMA) by the Congress of the United States in 1992, the U.S. Geological Survey (USGS) has been charged with building a national geologic map database. Under the provisions of the GMA, the Congress set forth the following objectives:

“The objectives of the geologic mapping program shall include-- (1) determining the Nation’s geologic framework through systematic development of geologic maps at scales appropriate to the geologic setting and the perceived applications, such maps to be contributed to the national geologic map database; (2) development of a complementary national geophysical-map data base, geochemical-map data base, and a geochronologic and paleontologic database that provide value-added descriptive and interpretive information to the geologic-map database; (3) application of cost-effective mapping techniques that assemble, produce, translate and disseminate geologic-map information and that render such information of greater application and benefit to the public; and (4) development of public awareness for the role and application of geologic-map information to the resolution of national issues of land use management.” The third phase of the response of the USGS to its charge from the Congress is the creation of an online geologic map database at a scale of 1:100,000. The National Cooperative Geologic Mapping Program (NCGMP) was established the Congress to be the responsible group within the USGS. The NCGMP then created the National Geologic Map Database Project (NGMDB) to actually design and build the national geologic map database.
**Goal of the NGMDB**

The primary goal of the NGMDB is a national geologic map database of which the third phase is a useable geologic map database that: 1) implements the North American Data Model Steering Committee’s (NADMSC) geologic map data model; 2) provides a seamless database of geologic map data; 3) provides versions of the data for publishing and investigative endeavors; and 4) provide web access and delivery of such data. Additional desirable properties are an ability to easily integrate geologic map data with other geoscience map databases, and capture better quality data during creation, editing, and maintenance of the database.

The first task of the NGMDB was to decide what constitutes the geologic map data to be assembled into a database. This decision was critical because the NGMDB recognized that the efficiency and capability to utilize geospatial data depends upon the structure and language used to organize diverse data sets, which the USGS is increasingly interested doing. This means that a robust data model needs to be developed for geologic map data in particular. The NGMDB has been working with national and international partners to improve and implement the evolving NADMSC data model.

**Uses for a Geologic Map Database**

The NGMDB recognizes several reasons to uses for a geologic map database. First, the use of geologic maps becomes easier in a digital environment. The easiest use to which multi-thematic maps can be put while in paper form is laying one map on top of another and then trying to see what correlations might exist. Any mathematical analysis from paper maps is quite time consuming and is usually not attempted. Analysis of the data from a geologic map extracted from a database is amenable to analysis by any number of software tools in such a rapid fashion that the process of analysis doesn’t intrude upon analysis itself. Geologic map creation can be greatly enhanced both in terms of speed, accuracy, and uniformity using a computer and when the compilation of a number of maps into a final map is the product, the time to produce a geologic map can be greatly reduced over tedious hand processes. Then publication at this point usually comes down to formats on the map page, including symbolization and choice of colors. Lastly, the exchange of geologic information including the ability to distribute knowledge of the content of the database increases the ability to use available data appropriately, and helps the USGS meet the mandate of Congress and the President for digital data and its general availability.

**The Nature of a Geologic Map**

**Descriptions and definitions**

Most of the data to be entered into the national database is in the form of published geologic maps. Realizing this, the next task involved discovering the nature of geologic maps in order to find a commonality from which a conceptual data model of paper maps could be constructed. Each piece of literature portrays another facet of the nature of a geologic map. Some examples will make this point clear. The American Geological
Institute’s “Glossary of Geology” defines a geologic map as: “A map on which is recorded geologic information such as the distribution and nature of rock units (the surficial deposits may or may not be mapped separately) and the occurrence of structural features (folds, faults, joints, etc.), mineral deposits, and fossil localities. It may indicate geologic structure by means of formational outcrop patterns and by conventional symbols giving the directions and amounts of dip, cleavage, etc., at certain points.” According to Maltman, 1990, “The geological map is a fundamental device of geologists … from which the geologist can deduce much about … (the) arrangement (of the rocks) underground and about their geological history… (and) can act as a synthesis of current knowledge on the geology of an area”. Compton, 1962, called, “Geologic mapping so essential to many field studies that it is sometimes considered synonymous with ‘field geology’”.

Oriel, 1974, said of geologic maps, “A geologic map may be defined as a summary graphic expression of available knowledge on the compositions, distributions and interrelations of earth materials at and beneath the earth’s surface as interpreted by a geologist for specified purposes on the basis of prevailing concepts”.

In a particularly insightful introduction to his book on geologic maps, Herbert Vossmerbäumer, 1983, says in the Introduction “Geologic maps are the “stuff of the trade” of earth scientists in academic and in applied areas. They are plainly the subject-specific form of documentation of geologic works (studies). They represent the consulting foundation for disciplines (that) depend on geologic knowledge, from agriculture to the raw-materials economy, building engineering to environmental protection.”

Some of the qualities of geologic mapping can be seen in the following quotes. Ridgeway, 1920, recognized a factor common to all mapping when he said, “It must be remembered that every map, whatever its scale, is a reduction from nature and consequently must be more or less generalized”. When ESRI, 2000, labeled the “Geologic Map of Iceland” as the “Best Cartographic Publication” at the 1998 ESRI International User Conference, ESRI was calling attention to the map esthetics and artistry common to most geologic maps (figure 1).

**Summary of Basic characteristics**

Some common ideas about what a geologic map is and what its components are can be extracted from the many statements above about geologic maps. Among these ideas are that geologic maps:

- Describe rock units – their compositions, ages, shapes, and relationships.
- Show structural features and their interrelationships with rock units.
- Show rock units and structural features as a consequence of their intersection with the topographic surface. This characteristic gives the geologic map a third dimension that is not generally present in other thematic map layers.
- Display the above concepts through the filters of the knowledge, interest, and interpretive skills of a geologist, the purpose for which the map was made (including the scale of the map), and the sense of map esthetics of the cartographer, usually the geologist.
The geological map of Iceland shows the main features of the bedrock geology. Formations are classified by age, type, and composition. The map also depicts the volcanic zones of the island and the distribution of the recent eruption sites. Lava fields of the Holocene are shown as prehistoric or historic.

Figure 1._ Geologic Map of Iceland
must contain report or story to explain the reasons for the geologic map and to explain the interpretations of the illustrated geology accompanying the map.

The taking of geologic information from a paper map including the accompanying report or from the immediate digital collection of geologic observations in the field and putting such data into a database requires the development of a data model that will give form and expression to the data in the database so that using the database would be nearly like using the original data sources directly.

**Concept of a Geologic Map**

Figure 2 describes a conceptual view of a geologic map that allows the translation of paper or observational map data to a database. The legend oval represents those geologic concepts displayed within the map. As examples the concepts contained in the map legend include formation names as well as the concept of a formation, fault names as well as the concept of faulting. The description oval contains descriptions of the concepts as well as of individual occurrences of the concepts captured as geospatial objects. Interpretations are also included in the descriptive material. The ovals then overlap to produce a geologic map.

**What is the data modeling process?**

**Creation of a Data Model**

Graham, 2001 says, “In general terms, a data model is a mathematical formalism for describing data and data structures (information) and a set of valid operations which are used to manipulate those data, or at least the tokens representing them. … Data models also help a designer eliminate redundancy in storage and inconsistencies in the structural design.” Four steps define the process that NADMSC and NGMDB have used while trying to learn how to build a geologic map database. These same writers also suggest that a requirements analysis, data modeling, and implementation sequence is an iterative process in which what is learned from previous attempts is used in succeeding data models and implementations.

Data model design can be broken into two parts: a requirements analysis and a conceptual data model. Modern database design authors recommend that database start with a requirements analysis. A requirements analysis helps to define critical concepts, their interactions, and other criteria. Next a conceptual model is created. This conceptual model helps to depict the world objects according to how the creator and/or user perceive them.

Database Design can also be broken into two parts: a logical model and a physical model. A logical model is a refinement of the conceptual model in the context of specific database system paradigm. Physical Modeling is the implementation of the Logical Model in a software system.

For source material for the requirements analysis the NGMDB has looked at:

Within the USGS

- National Cooperative Geologic Mapping Program: A La Carte a set of AML tools developed for ARC/INFO, GSMCAD an internal geologic map digitizing software
Figure 2._ Diagram of a Geologic Map
package, and the Southern California Area Mapping Project data model development effort.

- Mineral Resources Program which houses most of the databases in the USGS that are complements to the geologic map database.
- Evolving standards for contract digitizing of geologic maps.

External to the USGS, the NGMDB has partnered with, consulted with, studied or listened to:
- Geological Survey of Canada (GSC) – FieldLog, a digital methodology for creating geologic maps digitally and one of the prime designers of the current NADMSC data model.
- Petroleum Standards - POSC, PPDM

In addition, geologic mappers and geologic map producers have presented much material that has been useful in bringing the work of the NGMDB to its current stage at Digital Mapping Techniques (DMT) Workshops sponsored by NGMDB and Association of American State Geologists. These workshops have been held each year at different locations in the US since 1997.

**Conceptual Data Model**

The basic design goals for the conceptual model are built around a three-level data model (figure 3). The core part of the model includes the capture of geologic features and attributes that are common to most geologic maps. The core data include map units and structures, the map explanation or legend, and basic map symbolization. All of the data in the core part of the data model is data that the cooperators in the data model design process agreed is the minimal set of data that needs to be included in a geologic map database. This conceptual model framework allows for expansion of the data model. The area called defined extensions is meant to provide for an expanded set of features and attributes to be put into the geologic map database. Examples of geologic attribute data that might be included are overlay polygons of metamorphic grade, structural measurements that add to the knowledge of the stress field in the mapped region, engineering properties to describe properties of rock on which structures are to be built, and to describe properties of rock out of which structures can be built, and hydrogeologic information. The third part of the model, future extensions, might include data to help, as examples, land-use planners, biologists, and environmental scientists.

**Database Design**

Database design as practiced by the NGMDB for the purpose of creating a national geologic map database consists of two steps. The first step, logical modeling is the design, evolution, and refinement of the conceptual model in the context of specific database paradigm. The paradigm chosen for the design efforts that is reported in this paper is the object-relational paradigm. This is a system that allows spatial attribute data as well as non-spatial data to be represented as if the data were organized as if they were
Figure 3._ Conceptual Model Design
in an object-oriented model with the actual attribute data stored in a relational database. The NGMDB along with the NADMSC wants a database that is “useful”. Useful as applied to this discussion means that the database should:

- Use the NADMSC geologic map data model standard.
- Integrate with other geoscience databases. Analysis and comparisons with other databases is critical in insuring that the database is useful to prospective data users, but also to help verify geologic interpretations represented in the data.
- Be seamless - without information organized with data in a tile structure.
- Have the latest, most up-to-date information with alternative versions of the information available. This is important while making changes while editing, for correcting errors, making new interpretations, and for historical value.
- Make data quality checks easy to do and difficult to bypass.
- Web-accessible with a number of clients.

**Physical Modeling**

Early attempts by the NGMDB to implement the 1998 version of the NADMSC data model in a standard GIS package with a relational database were difficult to implement and did not completely satisfy the above criteria for a useful database. NGMDB immediately saw that to have a useful database supported by a robust data model an object-oriented data model would have to be created from first principals and then a new prototype database built.

Three reasons will explain the change in model paradigm to object-relational. First are the changes NGMDB has seen the thinking of geologists when they create geologic maps with traditional GIS packages. This causes geologic mappers to focus on two-dimensional geometry of polygons, lines and points first and what they represent second. Second, traditional GIS packages and database systems lack the capability to easily adapt to changing requirements of the data model and database. This is because the idea that since traditional GIS packages are built using a geometry-centered data model, attributes are stored in a relational database system that is separate from the geometry making the underlying data model hard to change without reloading the data (figure 4). Third, sheet-based or tile-based storage of maps common to traditional GIS software means that changes in one map that affect other maps in the same map library are not automatically changed at the same time (figure 5). The keeping of multiple versions of a map in such a system is a difficult bookkeeping problem, and web-based applications tend to need custom software that needs to be maintained. NGMDB has consistently wanted web distribution of the database integral to the development of the database to insure that web-based data distribution functioned as a built-in part of the design.

The choice of object-relational technology for the prototype database has the following desirable attributes. It is:

- More nearly a model of the real world
- Able to simplify tasks such as multiple representations
- A system in which data models are simpler and easier to build
- Less dependant on initial data model for future applications – the data model can evolve
- More flexible for meeting application needs and applications are simpler and faster to develop
Attributes are stored in a relational database system
Separate from the geometry

Figure 4a. Geometry-Centered Geospatial Data System

Fault - Real world object

Figure 4b. Object-Oriented Geospatial Data System
Figure 5a. Sheet or Tile Based Storage System

Figure 5b. Object-Oriented Storage System
• Easier to represent three-dimensional relationships
• Easier to implement complex rules such as: older rock units can be over younger units only if modified by tectonism, and dikes must be younger that the rocks they cut across
• Much easier to build a seamless database

Prototype Development

Proof of Concept

Implementation of the logical model for the series of prototype database efforts reported on here has been with object-relational GIS software package called Smallworld. The first prototype was in the terms of the NGMDB a “Proof of Concept”. It was built in 1999 using the relational data model as a basis for the data tables. Objects were then built on top of the relational tables so that the result would be close to what NGMDB and NADMSC wanted from a purely relational database without the necessity of large-scale changes in the then current NADMSC data model. This prototype was demonstrated at the annual Geological Society of America (GSA) in 1999 and the DMT meeting of 2000.

The Kentucky Prototype

The second prototype was designed with the cooperation and encouragement of the Kentucky Geological Survey (KGS). Colleagues from the GSC and the University of California at Santa Barbara (UCSB) took the lead in developing a new object-oriented data model to deal with database characteristics desired by the KGS. Geologic map data from four 1:24,000-scale geologic maps inside the Harrodsburg, KY 1:100,000 was used to build the database. This prototype had seven design goals:
1. Implement in the Smallworld system the revised conceptual data model that is an object-oriented version of the relational NADMSC data model draft standard. Add tables and attributes as needed to accommodate KGS data structure. Load the map information into the Smallworld system, and perform logical checks for spatial and geologic errors (e.g., stratigraphic nomenclature conflicts or mapping discontinuities at borders of adjoining data sets. Display the data with agreed upon symbology.
2. Develop capability for scale-dependent generalization, NGMDB recognizes that generalization and similar actions specified by user queries to the database are a long-term goal, but such user-specific, "interface" functions are not the priority of this prototype. Deliver appropriate level of map detail for user-specified zoom level for the current view of the database.
3. Demonstrate links from the map database in Smallworld to external data sets and databases. In this prototype, static versions of these data sets and databases will be used in place of dynamic links to the active data sets and databases. First, load in to the current database working session an ArcView shape file of hydrography, color it appropriately and be able to view the attributes of this data. Second, load the point coordinates for boreholes in the Kentucky coal, and oil and gas databases into Smallworld. Build and demonstrate the link to the external database. That is a user viewing the map database selects one of the displayed borehole points, and a pop-up display shows the coal or oil and gas data for that location. Tabular data,
images, and text files might be displayed from these databases. Third, load information for formal geologic units in the map database from the supplied URL for the relevant geologic unit name summary in the NGMDB GEOLEX database. Demonstrate the link allowing a user to select a map polygon, and then to see the relevant geologic name description in a pop-up window.

4. Cut out an arbitrary area clipped on-the-fly, as a user specifies an area.

5. Generalize rock units in the area selected in “4.” above by lithology and re-color the selected map units dynamically.

6. Demonstrate the ability to readily export the selected map data in “4.” and “5.” above without introducing spatial distortions (e.g., caused by map data re-projections) or loss of map data into Arc Shape format. Then deliver this data to the user over the Internet.

7. Develop a flexible approach to managing "stack-unit" map information, which can be applied both to the stack-units on the Kentucky GS maps and to the more complex maps that will be addressed in subsequent prototypes. As an example of more complex geologic settings, planned future work in cooperation with the Central Great Lakes Geologic Mapping Coalition will address the management of widely-distributed, thin and in places discontinuous, Quaternary deposits that commonly comprise a stack of 5-10 or more units overlying bedrock.

**Results of the Prototyping Experiments**

The NGMDB showed in the “proof of concept” prototype that:

- An object-relational GIS represented geologic features by scientific language rather than by geometric form,
- Editing changes to faults by adding rock units that cut a fault did not cut the fault into segments
- Geologic map data was delivered over the web in a form for a common GIS geologic client.

The Kentucky prototype was developed over fall and winter of 2000 and 2001, and was exhibited at DMT 2001 in Tuscaloosa, AL. The database was housed in Denver, CO, and accessed over the Internet using software from Citrix, Inc. Performance and speed of transmission were such that many thought that the database was on a local computer. NGMDB was able to show, in addition to the “proof of concept”, that for objective:

1. The revised object data model was implemented in an object-relational GIS package with relatively little modification to the data model. Both the data modelers and the database implementers learned how changes to the data model improved it and how careful use of the software allowed a fairly rigorous implementation of the data model. Map unit colors and line symbolizations were chosen by the KGS.

2. Display of the map data at different scales is a built-in function of Smallworld. Choosing colors based on chosen generalization of map units was also straightforward (figures 6 and 7).

3. The demonstration used a small version of the Kentucky oil and gas database for the four quadrangles in the database. The database is in Microsoft Access and was dynamically connected to the geology database by an ODBC interface. Values for the well symbol and captions were taken directly from the well database. Changes in
Figure 6. Overview Map with Outlines of Map Units in the Four-Quad Area.
Figure 7. Zooming into detailed area shows additional features, including map units coded by name, geologic contacts, faults, contours, geologic and economic locations, and streams.
these values obtained by changing the database were dynamically displayed on the current view of the geology database (figure 8).

4. 5. and 6. User-defined clipping of the database was implemented using three options: 1. Current view of the database, 2. Choice of quadrangle, and 3. Arbitrarily drawn polygon. The demonstrators first chose a well from the external database that was within 10 meters of a stream. An assumed leak from the well was able to move from the rock (limestone) in which the well was drilled, and from the stream, contaminate other limestone outcrops nearby. A network trace located limestone outcrops that were crossed by the stream and within a 2-kilometer radius of the starting well. The 2-kilometer radius circle was chosen as the clipping polygon. The rocks internal to the circle were generalized to lithology types and placed after this into a new version of the database so that it could be properly colored and common contracts removed (figures 9 and 10). The clipped data were then reformatted into ArcView shape files and exported to the local laptop computer in Alabama (figure 11). The database software interfaces well with a software package called “Feature Manipulation Engine”. The implementers exported clipped data set with little difficulty using this software package.

7. This item was not included in the prototype for lack of time. The concept of “stacked” geology and how this would be represented in the database has not been fully explored. However, this approach to especially Quaternary (surficial) geology is becoming increasingly important and if the geologic map database could be structured to hold data in this manner the database would help in Quaternary geologic studies.

CONCLUSIONS

The choice of an Object-Oriented Model

In choosing either a relational model or an object-oriented model for the base of the database design, the designer should remember that the origins of both paradigms lie in the world of business. Of the two the relational model is better known, has greater formalism, and is more familiar to database designers. Previous to this work most data models for geoscience data in the USGS were based on a relational model. However, more designers are choosing the object model. Gartner and Bergmann, 1999, say that, “… this way of modeling may lead to a closer approximation to real world conditions because of the consideration of the semantics of geo objects.” Worboys, 1995, says that “… geo data are characterized by a variety of attributes and behavioral values that define their spatial, temporal, and graphical dimensions as well as textural and numeric dimensions. The advantage of object-oriented modeling is the fact that objects encapsulate state (expressed by attributes) and behavior (specified by methods/operation) and they are able to communicate by sending and receiving messages.” In discussing the reasons why a user might wish to use a tool like DOOGIS, an object-oriented geo data program and geo data viewer from Argonne National Laboratory, the writers of the DOOGIS web pages said, “…What is an intelligent object? It’s an object that performs some behavior or supplies data, be that a simple road with address data, a tank driving across a battlefield, or a simulation of the Earth’s atmosphere. … we can dynamically couple to models, simulations, and/or applications producing behaviors and data.”
Figure 8._ A variety of external data may be displayed and used in the database without having to be part of the data model. One example is a set of oil wells that are maintained in a Microsoft Access database.
Figure 9. The database and external data may be used together to perform various analyses. For example, active wells can be examined to determine whether they have been spudded in a rock unit that is likely to form karsts, such as limestone. Furthermore, nearby streams can be traced to determine whether leakage could travel downstream to other karst-prone units.
Figure 10. The database can then be used to generate derivative maps based on primary lithology, for example, using quad boundaries, GIS view extents, or an arbitrary boundary at the extent.
Figure 11. The derivative map may be exported from the central database to a LOCAL desktop in other GIS formats for further analysis.
The NGMDB has found by experiment that most of the promise of object-oriented data models is true.

Problems with Object-oriented Models

A number of model designers have suggested that object-oriented models are not without fundamental problems. Graham, 2000 propounds several problems with the technology, a few of which are:

- “Most object-oriented programming languages do not support the notion of persistent objects, that is objects which, stored on disk, persist unchanged between executions (of software)”. This means that programming databases in current o-o languages is difficult at best.
- “… sometimes it becomes necessary to trace the behavior of (an o-o) system, in order to debug it, by tracing the passage of messages from one object to another … (and) can be enormously difficult and remains, in my view, a very serious defect of object-oriented systems.”
- “… no system in wide commercial use supports class and instance level inheritance of methods, attributes and attributes with explicit, user-defined rules for conflict resolution.”

Muller, 1999, says that when implementing an object-oriented data model in an object-relational system, “Without an accepted standard, … you have to base your approach on the quirky tools you have at hand.”

Miah, 1997, writes an extended critique of the object-oriented paradigm. Very briefly, he says:

- “(The object-oriented paradigm) 1. Assumes that the world is made of objects and relationships, 2. Assumes human beings think in terms of object and classes.”
- “In real life, some classes are more typical than others (e.g. robin vs. emu as typical bird); if has the idea of class the (data model) should recognize this.”

In the data model used in the latest prototype, an object class “boundaries” contains faults and contacts. Where do faults that are contacts really fit in such a model? Are fault-contacts to be placed in their own object? How are atypical instances of objects handled?

**Work to be done**

Science language, the accepted word lists, are far from complete. The categorization of rock types and structures, to name two areas, has just begun within the NGMDB effort. Also, much investigation of object-relational technology needs to be done to see how well the data model under development by the NADMSC can be implemented in such technology. Further, the prototypes have, so far, investigated the ability of the current data model and software system to handle geologic map data over a narrow range of geologic environments. The NGMDB and the NADMSC fully expects that major changes to the current data model and that much work is ahead will be needed to fully encompass the range of geologic map data for a national geologic map database.
References Cited


Müller, Robert J. Muller, 1962, Database design for smarties: Morgan-Kaufmann, 442p.


