GEOLOGIC DATA MODELS AND CORDILLERAN GEOLOGY

Submitted for GSA Memoir (Topical Session 85)

Andrew Zolnai
ESRI Inc.
380 New York Street
Redlands CA 92373
azolnai@esri.com

Abstract

Two facets of field data management will be addressed as relates to Cordilleran geology – data collection (field data, data models, object database and refinements), and data models (accuracy, breadth, commonality and depth). This will illustrate how accuracy in geological mapping and interpretation, need no longer by a compromise of computer system speed and storage capacity. New taxonomies also help share and disseminate geological data and interpretations, so important to state and federal surveys worldwide.

1. Introduction

Geological field data capture is documented by the British and other geological surveys (www.bgs.ac.uk/dfdc). Data models are maintained by the US and other geological surveys (http://geology.usgs.gov/dm). Brodaric et al. posted a geologic data model (http://support.esri.com/datamodels > geology) that proposes to separate 'concept' from 'occurrence' data in the context of a geographic information system (GIS). Concepts are geologic features such as contacts (stratigraphic, unconformity, fault, etc.), classification (lithology, age, etc.), etc. Occurrences are recorded in the field (points, lines or polygons depicting concepts). Distinct sets of attributes help maintain fidelity and accuracy: field locations can be modified regularly (by field season or by team), but the underlying concepts a lot more rarely (say, a rock classification is proposed or modified). If properly set up, the one can be reflected in the other automatically, without touching the counterpart attributes.

Database storage also allows all to benefit from software advances over time. Scale dependency for example helps to control the display of data - details such as strike and dip above 1:24,000 scale against satellite imager under 1:125,000, or digital elevation data at sub-meter accuracy against 150 m resolution at those same respective scale ranges. And further refinements can be built into the database itself: the petroleum and mining industries evolved data modeling taxonomies, the USGS to rank internet petroleum data searches by area (http://seamless.usgs.gov), and mining applications for ranking and interpretation of ore deposits (www.georeferenceonline.com). Let us look at some immediate benefits for Cordilleran geologists.

2. Geologic Data Models

Cordilleras offer the benefit and challenge of topographic complexity, both in geomorphology (topography, land cover) and geology (tectonics, stratigraphy, time). From pioneering regional work (Douglas et al. in Price, 1994) to recent interpretations (McMechan, 2001), accurate geologic models have been matched by the sharing and dissemination of same (http://hub.geosemantica.net). The US Geospatial One-Stop (www.geodata.gov) or UK Digital Energy Atlas Library (www.ukdeal.co.uk) are examples of increased public access in both directions (download and sometimes upload protocols). Private industry and public agencies alike see increasing demands from their user and client base as a result.

One example is increasing transparency in current practices (Browne, 2002). It necessitates a language that speaks not only to the industry itself (earth scientists here), but also to the public users and client agencies at large. The North American Data Model mentioned earlier, attempts to articulate complex geologic concepts in database format, chiefly to help state and federal agencies communicate among themselves. The GEON project also assembles the geologic maps for the SW USA on the web (www.geongrid.org), but their remains still the work of matching nomenclatures at state boundaries. This is where the "A-to-D"s of data interoperability can help:

a. accuracy

It has been noted in structural geology that small scale structures (cleavage, slickensides, etc.) often reflect larger ones (faults and folds, etc.). Data recorded at various scales can be preserved in databases that scale not only in capacity (memory and speed) but also in accuracy. This is done in two ways: Data models or templates allow to specify ranges and precision that are appropriate to various scales as mentioned earlier (this can be done in both relational and object databases). User groups defined a data modeling frameworks to meet their needs (www.gita.org/events/oil_gas/03/program.html > Standards and Metadata).

Secondly, workflows allow to document and repeat the procedures once they have been established, or as they are being refined iteratively over time and/or by various teams. Both processes afford ways and means to set up data ranges and precision as appropriate – field data points collected with GPS will afford sub-metric accuracy, survey station series in the metric range, and satellite images in the tens of meters, for example.

b. breadth

The current challenge is to render digitally the wealth of information compiled for example in Decade of North American Geology. The USGS' National Atlas has a geologic derivative (http://tapestry.usgs.gov) – but how can the depth and breadth of data be bridged? The link is with data dictionaries, metadata and best practices (www.gita.org/events/oil_gas/04/schedule.html >

Standards and Metadata Part II). This allows various differing datasets to be linked in a meaningful manner, without having to recreate the same according to rigid rules.

Metadata have previously been mandated at US federal agencies (the so-called FGDC standard). State agencies, industry or academia on the other hand have focused on documenting best practices, and converting them into workflows that can be shared, repeated and refined – users find what makes sense in their daily routines, store the relevant information as metadata, and document the work processes involved. Object databases also provide tools to store metadata and record workflows (www.esri.com/opengis). The aim is to render explicit what is implicit in earth scientists' minds, in order to translate geologic data into machine language and share it with others.

c. commonality

If metadata and workflows match user needs, then the basics are in place to set up a data dictionary framework. No two agencies will have the same base data, needs or mandates – it is more practical to link what they already have, than to (retro)fit it all into a comprehensive (and elusive) master database. Take the very simple case of a legend: it is most commonly sorted by age, and lithologies or formations are classified (and repeated) underneath; this makes for a rather exhaustive column that increases exponentially in size with increasing age and geographic range. And if this worked for local experts, this was hard to convey meaningfully to geologists or the public not familiar with each area.

The geologic data model can help again – it details explicitly what specialists have painstakingly recorded over time. That documentation can in turn be included in data dictionaries, and used to link other maps sheets. Both object and relational databases can include rules and behaviors that specify how to link data sets. Topology can for example be used to tell that a fault can cut a contact but not the reverse. Rules for example can depict how formations are grouped into sequences, and thence into time-zones. Only GIS can however do this in the spatial domain, which is absolutely critical to field-mapping.

d. depth

As mentioned earlier, Cordilleran geologists have the opportunity and the challenge to depict and interpret complex geology in three physical dimensions plus time. This requires comprehensive and accurate data for modeling. Computer systems have been around for a while (Pflug and Harbaugh, 1992), but have tended to be point-solutions, well suited for specific tasks, say in 3- or 4-D modeling, but not for interoperability or sharing (Zolnai, 1992). The latter is however paramount for the development of complex models among teams of specialists. Consortia have again addressed this – gOcad for example was funded by a number of companies including petroleum (Mallet, 1991), and their 3D tessellation object model has been incorporated in oil and gas applications.

Such systems have not however remained open in the modern sense – adopting IT standards in programming (VB, C++ etc.), and systems (.NET, Java etc.) and protocols (SOAP, XMI etc.) that ultimately allow the sharing and spreading of models. Adding 3D modeling features is an integral part of recent software developments depicted above – in other words, a careful and methodical approach toward open and interoperable 3D systems will be based on data models and metadata discussed above. Figure 1 shows the demo dataset posted with the geologic data model discussed above, fully accessible in a 3D system. This will lead to future open 3- or 4-D modeling based on the criteria listed above.

3. New Paradigm

Earth scientists are now at a place where accuracy and speed no longer are trade-offs against one another according to available computer performance – system speed and storage capacity are no longer a constraint. Accurate data can be stored in large and/or distributed systems. Detailed data models can be made to meet user needs. Metadata and dictionaries allow to share both data and models among professionals as workflows and best practices. The same can be shared with the public and investors on the internet and help disseminate the acquired knowledge. And such systems are the foundation for further 3- and 4-D modeling of geologic systems needed for Cordilleran geologists.

This is an interim report, as full 3- or 4-D modeling is not described here. The foundations are still being built, but readers are offered a glimpse of tools to come, that will truly help Cordilleran geology evolve rapidly over time. The key point is this: if openness and interoperability are complex in IT systems, they are even more complex in geologic modeling. But data models, metadata and dictionaries, best practices and workflows will stand earth scientists in good stead. This is currently the case of other physical sciences such as atmospheric modeling (www.rap.ucar.edu/weather). A whole new language is in the process of being built in order to study and predict complex systems in an open a sharable environment.

4. The Road Ahead

Two facets of data management and interpretations were covered so far: data collection (Field data, data models, object databases and futures), and data models per se (accuracy, breadth, commonality and depth). This will help translate current processes from the analog (notes, maps) to the digital (data bases and models). Work flows and best practices based on metadata and dictionaries, will render day-to-day efforts by earth scientists repeatable and sharable over distributed systems. Accuracy and consistency can thus be properly addressed and controlled via processes that are (“three S”s):

a. scalable - record and process data at scales and accuracies that are appropriate to each feature, on systems that are appropriate to the amount of data available

b. stable - create data models and metadata where new fieldwork or evolving concepts don't affect each other, and on platforms that can evolve and change over time

c. shared - store and manage data and interpretation as processes that can be upgraded an maintained as work progresses on systems that evolve, among groups of specialists and the public alike

Cordilleran geologists also have significant needs in 3- and 4D, in complex tectono-stratrigraphic domains. Data management and modeling building blocks described above will help this complex process in addressing accuracy and repeatability issues. GIS offers a scalable and interoperable vehicle for geologic modeling that is currently being extended into 3- and 4-D. This will also fit in a global infrastructure of shared data and resources on intranets and internets alike, and allow a dialogue among earth scientists, the public and investors alike.

The end-result will be a truly Global Spatial Data Infrastructure, like that designed for various agencies worldwide (www.esri.com/gsdi). This is part-and-parcel of Societal GIS, which aims at creating a common language to help manage our Earth's resources. Ray Price pointed out fifteen years ago how Geoscience contends with global change (Price, 1989). As earth scientists in general and cordilleran geologists in particular, it behooves us to equip ourselves with the best descriptive and predictive tools as custodians of our Earth – only then can we help others model and remediate processes we Humans initiated recently in geologic terms.

References

Browne, J., 2002: Transparency key to restoring trust, Oil&Gas Journal, 28 Oct. 2002

Mallet, J.-L., 1991: GOCAD a computer-aided design program for geological applications. In: Turner A K (ed.) Three Dimensional Modeling with Geoscientific Information. Kluwer, Dortrecht (see also: http://www.32igc.org/abstracts/T1002.pdf)

McMechan, M. E., 2001: Large-scale duplex structures in the McConnell thrust sheet, Rocky Mountains, Southwest Alberta, Bulletin of Canadian Petroleum Geology, Sept. 2001, p. 408-425

Pflug, R. and Harbaugh, J. (eds.), 1992: Computer Graphics in Geology - Three Dimensional Computer Graphics in Modeling Geologic Structures and Simulating Geologic Processes, Lecture Notes in Geology 41, Springer-Verlag, New York

Price, R. A., 1994: Cordilleran tectonics and the evolution of the Western Canada sedimentary basin. In: Geological Atlas of Western Canada. Mossop, G.D. and Shetsen, I. (eds.). Calgary, Canadian Society of Petroleum Geologists/Alberta Research Council, p. 13-24.

Price, R.A., 1989: Geoscience contends with global change, Comment, Geotimes, p. 9, March 1989

Zolnai, A. and Marchand, M., 1992: Beyond GIS/GPS – Trends in Spatial Data Handling, GIS World, Nov. 1992, p. 52-54

Figure 1

Toad River sheet from the geologic data model, displayed in ESRI 3D Analyst extension (ArcGlobe), with elevation data from Canadian Council on Geomatics (www.geobase.ca)