Improving the Efficiency in the Maintenance of the Provincial Geological Database

by Y. Cui

KEYWORDS: geological map, spatial database, maintenance, update, GIS, MapPlace, best practices

INTRODUCTION

This paper describes some of the best practices that are promoted by the British Columbia Geological Survey (BCGS) to improve the efficiency in maintaining geological maps. The main focus is to leverage interoperable (including freely available) spatial database technology to reduce the redundancy and efforts in map compilation and data integration while enhancing the data quality.

BCGS is responsible for producing and publishing province-wide geological maps through MapPlace (http://www.empr.gov.bc.ca/Mining/Geoscience/MapPlace/Pages/default.aspx) to mineral exploration, mining, land planning and other users. BCGS recognizes the value of a corporate database management environment for a seamless digital data flow from outcrop to dynamic maps on the web.

The most recent province-wide geological map was published at a scale of 1:250 000 scale in 2005 (Massey *et al.*, 2005). The map compilation was made in support of the Mineral Potential Project (1992-96), with updates in 2003 and 2004. The province-wide geological map has not been updated since 2005, except certain focused areas such as QUEST (Logan *et al.*, 2009).

Currently, new map compilation and updating are carried out by BCGS staff members using Manifold® System and MapInfo® desktop GIS tools. Data files are distributed across different file servers and desktop workstations of the mapping geologists. When a mapping project is completed, a version of the data is converted to the appropriate format and made available to the public through MapPlace. In addition to publications such as Geological Fieldwork, Open File, GeoFile and Geoscience Maps, geological maps are also available online as PDF files and GIS formats for download from MapPlace.

The GIS software tools (*e.g.*, Manifold® and MapInfo®) are sufficient in map compilation and analysis when working with a relatively small dataset and a limited number of data layers. However, this environment has a number of challenges and limitations, including:

- difficult to apply common data standards among the mapping geologists;
- unable to manage and apply data quality rules;
- difficult to share consistent legend and map styles;
- difficult to manage multiple versions of the same map at varying stages of editing and correction;
- slow performance in loading and incremental saving of edits for a large map; potential risk of losing un-saved work when the system crashes (more frequent on workstations with certain operating systems);
- limitation on accessing and loading a large volume of data, *e.g.*, province-wide topographic base map which is required to node the geological features;
- unable or difficult to collaborate amongst staff members on compiling adjacent mapping areas, resulting in discrepancies along the map boundaries and inconsistencies between mapping areas;
- unable to automatically create derived products (*e.g.*, different scales and customization) from the source geoscience data;
- difficult to manage multiple versions of the same map at different scales and customization;
- no automation in detecting and correcting defects or inconsistencies in the geoscience data; and
- difficult or unable to discover and access datasets from different sources, in different formats and projections, and stored on different locations, including staff's workstations.

The geology operational database environment (GODE) is proposed as part of a solution to address some of these issues at BCGS. This paper briefly describes the architecture of GODE, with a focus on the recommended best practices to maintain the operational database.

Typically, when organizations change operational environment, they should consider not only the specific

This publication is also available, free of charge, as colour digital files in Adobe Acrobat® PDF format from the BC Ministry of Forests, Mines and Lands website at http://www.empr.gov.bc.ca/Mining/Geoscience/PublicationsCatalogue/Fieldwork.

business drivers and user requirements, but also the existing processes, tools and most importantly the culture and human aspect in adopting changes. At BCGS, the following requirements are addressed in order to document a high level view of the GODE system architecture and design:

- full context of relevant map layers is available and not constrained by the volume of data: base maps, existing geological maps, new updates from other mapping projects, archived and retired or historical maps;
- base maps that are readily available and at the appropriate scales with consistent styles for new map compilation or updating, to provide the geographic context and geo-referencing in some cases;
- adopting consistent data model that can be enforced by schema validation and schema mapping;
- a custom user interface to Manifold® and MapInfo® for data capture and manual data entry to improve efficiency and to reduce errors;
- a workspace where the province-wide geoscience data, including other project geologists' maps, is readily accessible for data quality assurance and integrations (*e.g.*, boundary issues, connectivity, currency);
- an environment to support versioned map compilation and data archiving;
- reduced inconsistency in data through database triggers and constraints based on data access policies and data quality rules;
- shared legends and styles: if a new rock type and associated styles are defined, they become available instantly to all the users the moment they are implemented in the database;
- ability to carry out province-wide statistical and spatial analyses: *e.g.*, spatial overlay of bedrocks, faults, geochemical results, geophysical survey results;
- ability to create derived products through province-wide data processing;
- ability to provide up-to-date information through MapPlace directly served up from a database; and
- provision of dynamic provincial geological map as GeoWeb services such as WMS, WFS and KML.

DEVELOPMENT OF OPERATIONAL ENVIRONMENT

Guiding Principles

The development of GODE and the recommended best practices follow a set of guiding principles that are common in service-oriented architecture:

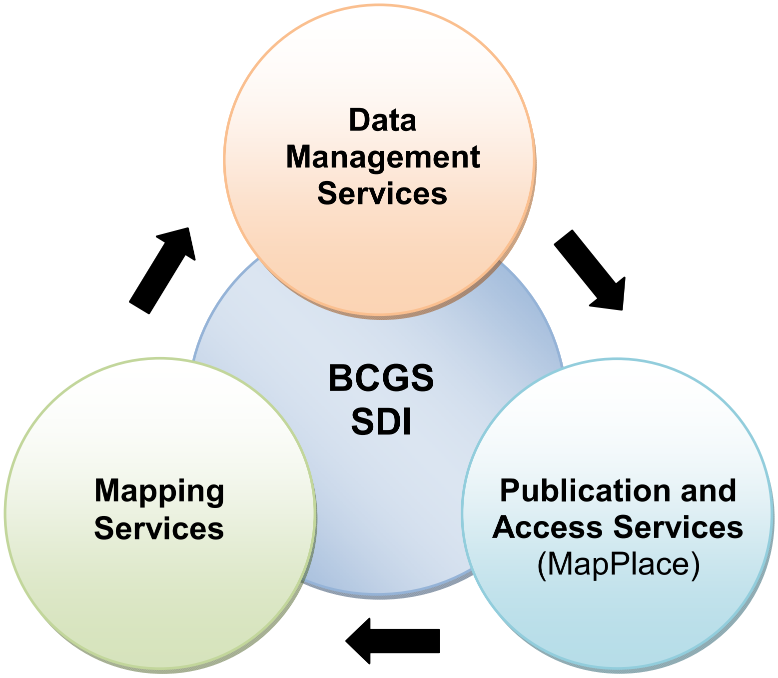
- **Interoperable**: GODE will be designed and developed based on ISO and open standards and specifications to ensure meaningful interoperability with other systems and services.
- **Scalable**: GODE will be designed to manage cumulative new map compilations and versioning of the integrated province-wide geological map.
- **Encapsulated**: Map maintenance users should be able to access all the relevant data and use the tools within GODE without the need to be exposed to the complexities of the underlying technology and information infrastructure.
- **Collaborative**: GODE will be developed with a "locking" mechanism and user profiles to facilitate collaborative map updating, especially at map boundaries of adjacent project areas.
- **Efficient**: Data quality and validation rules will be formally specified and automatically applied wherever possible to eliminate simple defects in the data and reduce repeated efforts, to sustain long-term maintainability.

Business Drivers

The requirements of GODE are largely driven by the business needs and processing cycles at BCGS. The collection and management of geoscience data are logically related to three service components and can be functionally grouped and deployed on a common Spatial Data Infrastructure (Figure 1):

1) Mapping Services: geological field mapping, mineral potential assessment, geochemical survey, and geophysical survey;
2) Data Management Services: map compilation, data quality assurance, data integration, and production; and
3) Publication and Access Services: MapPlace and other third party GeoWeb services.

Publication and Access Services are the front end where products and services are provided to clients. Access Services are positioned to collect and communicate user needs, which are documented as updated requirements and specifications of products and services. This would drive Mapping Services with mapping projects to address the new and changing business needs. Updates and new mapping results are available to Data Management Services to produce



products and services that Publication and Access Services rely on.

System Context

GODE forms part of the Spatial Data Infrastructure (SDI) at BCGS to facilitate the management and provision of geoscience data. The SDI provides a suite of client-facing GeoWeb services enabled by standards, spatial database, and software components. Figure 2 shows the high-level context view of BCGS GeoWeb services and how the components or services interact with each other. The modelling and system architecture of the SDI at BCGS loosely follow ISO/ITU standard 10746-3 on Reference Model for Open Distributed Processing (*e.g.*, Farooqui *et al.*, 1996; ISO, 1998; ISO, 2010) and some of the recent work on modelling SDI (*e.g.*, Brodeur, 2011; Cooper, 2007; Hjelmager, 2005; Hjelmager, 2008).

System Architecture

To briefly describe the system architecture of GODE at a high-level, graphical depictions of a few selected viewpoints are presented here to illustrate the system domains, components in each domain, actors and interfaces. The enterprise viewpoint describes the purpose, scope and policies of the SDI. Figure 3 shows the actors (GODE Operator), domains (*e.g.*, Data Sources, GODE, and Geology Application Database Environment), and components in each domain in the context of this enterpriser viewpoint. The information viewpoint consists of BCGS geological data models enhanced with additional metadata to manage versioning and keep track of the various stages of initial field observation, quality assurance, integration, archiving and production. Due to its large volume, details of the information viewpoint are not included in this paper. The computational viewpoint (Figure 4) depicts the decomposed database components and interfaces that are required for the GODE to function. The engineering viewpoint (Figure 5) shows how the components used in the system are distributed across the various servers.

The technology viewpoint describes the hardware, software versions and other technologies used in the system. The feasibility study and prototyping of GODE rely heavily on free and open source software, including PostgreSQL/PostGIS as spatial database, OpenJUMP as desktop GIS, JEQL as a query and batch Web processing engine, and Apache PHP running on Windows Server® 2003 and Windows XP® and Vista® on workstations. The test deployment also includes Microsoft® SQL Server® 2008 R2 running on Windows Server® 2008 R2 64-bit, and Manifold® System as desktop GIS running on Windows Vista® 32-bit workstations.

Open distributed processing amongst the various system components and servers that are part of or related to GODE, is shown in Figure 5. These are supported by the adoption of Open Geospatial Consortium (OGC) standards, including the OGC Simple Features Specification for SQL (OGC, 1999). Spatial data, encoded as GEOMTRY or GEOGRAPHY data types on the database side, can be exposed as well-known text (WKT) or well-known binary (WKB) (OGC, 2001; OGC, 2010), to desktop clients and web services. This ensures

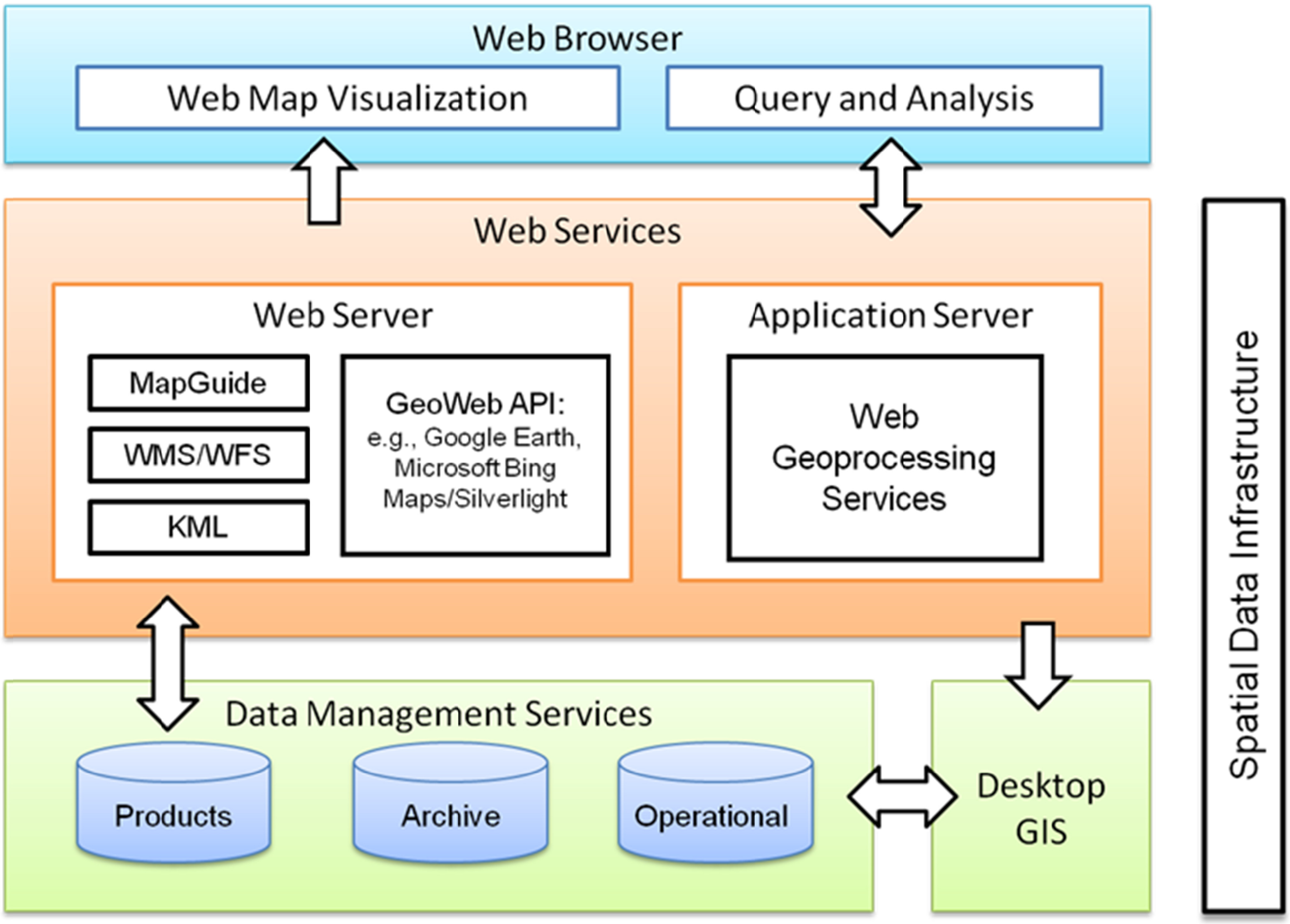


Figure 2. High level context view of BCGS GeoWeb services.

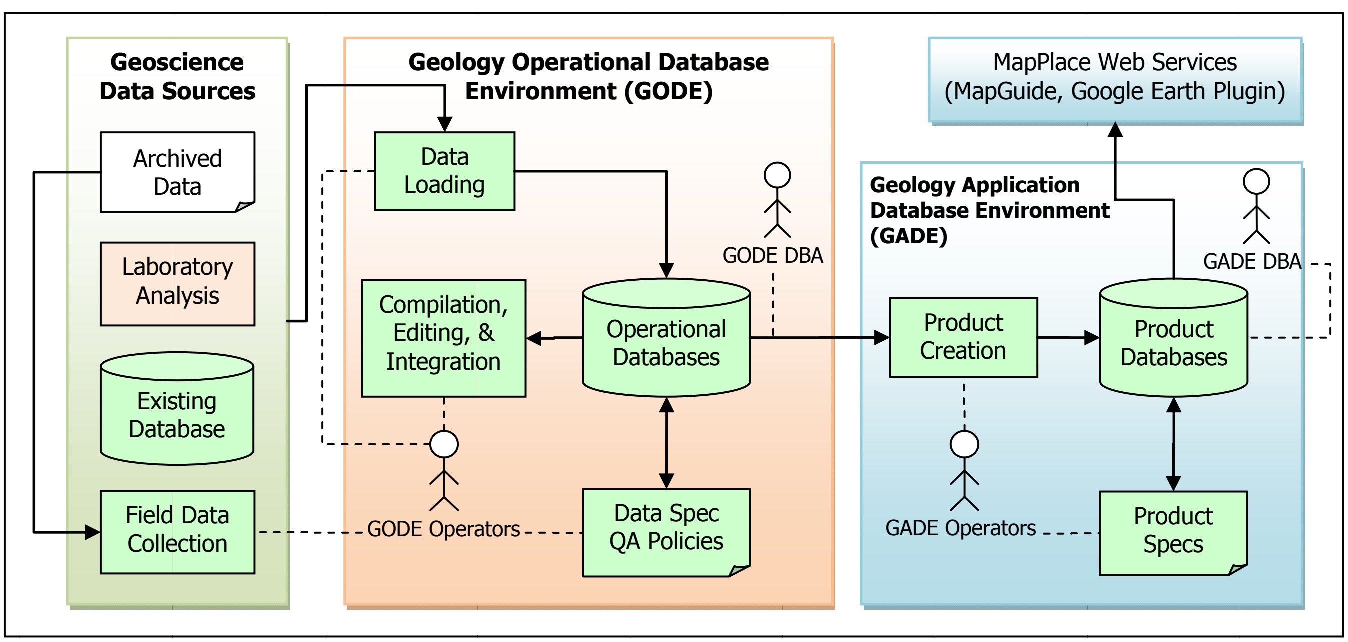


Figure 3. Enterprise viewpoint of GODE within the BCGS Spatial Data Infrastructure (SDI).

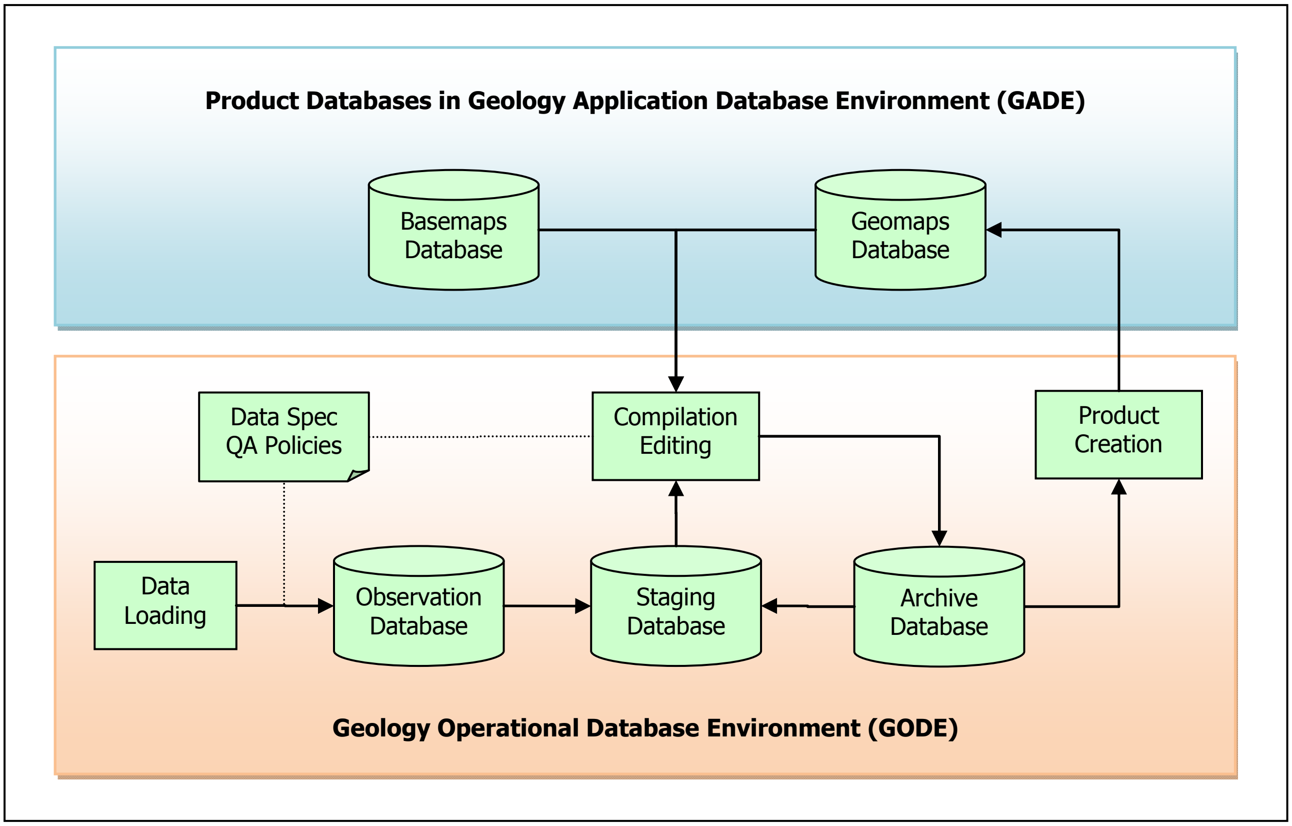


Figure 4. GODE database components.

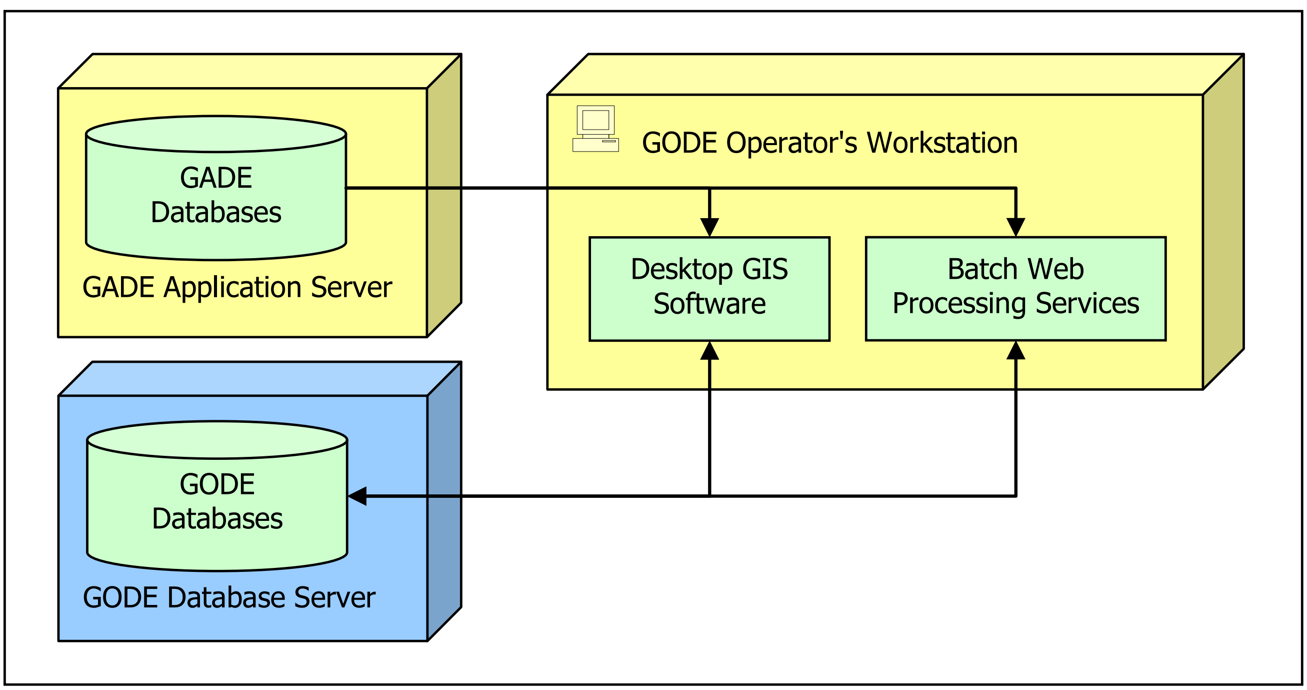


Figure 5. GODE engineering viewpoint.

interoperability at a more primitive and practical level without the need to parse a markup language with a complex schema and overloaded with redundancy.

BEST PRACTICES

Data Quality Assurance Rules and Policies

Geological mapping has multiple stages, from field survey, map compilation, cartographic enhancement, to final production. In addition to rigorous training and years of practical experience by the seasoned mapping geologists, there is also abundant literature on the geological mapping specifications and techniques for the various map producing stages, including the BC RIC standard (RIC, 1997), Ontario Digital Line Standards (Muir *et al.*, 2000), unpublished guide and manuals at BCGS, and over a decade of "Digital Mapping Techniques" workshop proceedings from the Association of American State Geologists and USGS (http://ngmdb.usgs.gov/Info/standards/datacapt/datacaptureWG.html).

The recommended best practices are derived from these established and well-adopted standards and guidelines with an emphasis on achieving efficiency in the configuration of map compilation environment and the specification and applications of data quality rules, such as:

- size of features: minimum area size of polygonal features, minimum length of linear features;
- density of coordinates: too many coordinates or too few coordinates; and
- geometric irregularity: duplicates, kick-backs, sharp angles between two lines, overlaps, overshoots, undershoots and gaps.

Map Projection

At different stages of geological mapping, a specific map projection is preferred to meet the requirements. In the initial map compilation, it is essential to ensure the proper positioning of the field observations with referencing to the most detailed topographic map, to preserve measurements of angles and orientation of directed linear features (*e.g.*, thrust fault traces). As an example, the choice of projection is conformal while the use of UTM is preferred by many geologists because orientation data works better with the orthogonal grid. At the stage of province-wide map integration and map production for applications such as statistical analysis, the choice of projection is equal-area based. In British Columbia, the preferred projection is BC Albers.

It is well known in the geospatial user community that coordinates can drift and shapes can change after map projection and round-trips of re-projections (*e.g.*, from geographic coordinate system WGS84, to BC Albers, to UTM and then back to WGS84). While it is impossible to avoid projections, there are measures that can be taken to reduce the coordinate drifting, including proper use of explicit unit of precision, coordinates densification for large features that do not have enough coordinates along straight and long edges, and avoiding clip or cookie-cut of contact lines when checking out the features for a mapping project area (more in the section of "Checking-out" Existing Geological Maps).

For maps published in a non-conformal projection, orientation is distorted in areas not at the central meridian. Certain orientation data (*e.g.*, the strike of a bedding) can be stored in a database as true north and displayed as symbols. This stored orientation data should be projected so that it is displayed consistently with the "distorted" orientation of other geological features.

Awareness of Varying Mapping Scales

Traditional systematic geological mapping at a fixed scale for a map sheet has been replaced by mapping at varying scales designed to answer specific geological questions or targeting economic mineral potential. At BCGS, project-based mapping can be carried out at a scale of 1:10 000 to 1:50 000, and the publication at regional or provincial extent might be generalized at a scale of 1:250 000 to 1:2 000 000.

It is worthwhile to include the mapping scale in the metadata and adjust the data capturing with positional accuracy, unit of precision (see next section), and level of details appropriate to the map scale. Digital mapping provides a perfect breeding ground for imperfect mixing of data captured at different scales in the same area or maps at different scales adjacent to each other. Data processing with scale awareness can treat the data at finer granularity while performing data validation and data quality assurance.

Explicit Unit of Precision

Most GIS and database systems can store coordinates and perform computation at a unit of precision below microscopic resolution. It is perfectly fine to use the highest precision that a system supports for computation. However, it can cause data quality issues with an excessively finer precision or different units of precision for data stored in different systems or formats, for mapping at a scale of 1:50 000 or smaller.

In GODE, map compilation is to be carried out at decimetre precision (or 7 floating points in decimal degrees), and the data is maintained at metre precision (or 6 floating points in decimal degrees) on the database side.

Common Topographic Base Maps

Topographic base maps are essential components for geological mapping, as background and cartographic enhancement to the final publication, as one of the sources for georeferencing or checking positional uncertainty, and also as noding bases to close off certain geological boundaries to the topographical boundaries such as lake shores, river banks and coastal lines.

Corporate spatial database makes it possible to store a very large volume of the most detailed topographic base maps and to retrieve any given map sheets easily.

Another major advantage is that the styles applied to the base maps can be stored on the database side and shared by every user. This approach not only saves time on discovering, retrieving and styling base maps, it also provides visually consistent base map layers to the final publications.

Provincial orthophotography and other digital images at high resolution are available as WMS layers or image files to validate the accuracy of important geological features.

"Checking-out" Existing Geological Maps

Geological map updating can start during field surveys, which would require the "checking-out" of the existing geological map from the corporate database for the project or target area. A common practice is to use map sheet or project area neatlines as a cookie-cutter to clip out only the portion of the map in the area. It has been proven that this practice can cause lot of data issues by the time the updates are to be integrated to the corporate database.

It is highly recommended that the project area neatlines are used to intersect all the polygonal features (*e.g.*, bedrock polygons) in their entirety and then use the polygons to extract all linear geological contacts and faults that either form or intersect these polygons. When the updating is completed, the integration process should only involve accepting features marked as one of the following:

- no change,
- new compilation,
- revised with updates: geometries, attributes, both geometries and attributes, or due to integration), and
- retired (*e.g.*, replaced by updates).

Attribution through Standard Lexicon Templates

It is a challenge to standardize the nomenclature of geological units for an area as large as British Columbia with diverse geology and mapping contributions, primarily by BCGS, but also by federal government agencies, universities and mining industry. Nevertheless, an attempt is to derive a standard template by reconciling the differences between the Canadian Geoscience Knowledge Network (CGKN) lexicon entries for British Columbia and the ones existed in the current Geological Map of British Columbia database.

In order to accommodate more detailed mapping and subdivision of existing major geological units, a template for sub-units is also created.

During initial map compilation, the geological unit code mapped to the standardized lexicon in the template should be used to simplify the population of the attribution. Whereas the geological units are subdivisions of an existing major unit, then a template for the sub-units can be used or new sub-units can be created.

Map Neatline (Knowledge Boundaries)

A province-wide geological database could contain project-based mapping updates at multiple scales and different stages of completion. This would leave inconsistency, lost connectivity and other issues at map boundaries. While it is possible to produce a product at the smallest scale as a common denominator, the inconsistency and boundary issues will remain to be resolved at the observation level and has to be managed properly. One recommendation is to introduce a new contact type, "neatline", to separate areas of different mapping scales or stages of updating. In that sense, neatline is also the knowledge boundary, meaning a lack of knowledge occurs beyond the neatline, at the level of detail that a recent mapping project has taken place. The neatline is removed only when knowledge becomes available to resolve the inconsistency, connectivity and other data issues at the map boundary.

Treatment of Small Geological Units

There are bedrock units that are geologically or economically important, as determined by the mapping geologist, but they are small in size or spatial extent. These features should be kept in a separate map layer or a separate table in a spatial database. While most of them are mapped as polygons, some of them are best digitized as points or lines such that they can be properly symbolized or styled for display.

Mapping Mineralization

Mineralization or mineral occurrences can be mapped as points, lines and polygons. In most cases they are captured in a separate map layer or maintained in a separate table in a spatial database. In some cases the mineralization can be annotated through the use of a modifier code on the geological unit designation, preferably as a sub-unit, with explanation in the description.

There are also situations where the mineralization is significant both in importance and geographic extent and the hostrock or original rock types cannot be reliably traced through the zone of modification. In this case, it is acceptable that a new geological unit is created.

DATA UPDATING AND INTEGRATION

Data Loading

When a project-based map is available for updating the province-wide geological database, the contact lines (including regular contacts, faulted contacts and faults) and bedrock polygons are loaded into the observation database after schema mapping and validating projections, scales, and geo-referencing. The data remains in its original state without any changes or fixing.

A mandatory column in the data is a lexicon-based geological unit or unit code which must exist and can be matched to the lexicon major unit or sub-unit templates. If it is a new unit or subunit, full attribution must be provided.

Noding and Polygonization

A copy of the contact lines and existing province-wide geological contact lines are loaded into the staging database (Figure 4) for detecting changes, adding new features, replacing or retiring updated features, and noding contact lines. The resulting contact lines are used to form bedrock polygons.

Both the noding and polygonization can be carried out on the staging database side, through desktop GIS or batched processing services. At BCGS, most of the work is carried out in Manifold® System with linked in maps (or database tables) from the staging database.

Data Quality Assurance

During or after the noding and polygonization processes, a number of data quality assurance (QA) procedures are repeated to check and fix some of the potential data quality issues against the data quality rules. The QA rules and policies can be specified and stored on the database side to ensure the QA rules are not only enforced but also applied consistently. Another benefit is the potential to apply the QA rules automatically through the development of database side triggers, constraints, SQL scripts or other means accessing the same QA rules and policies stored on the database side.

At BCGS, the current QA work is carried out in a hybrid approach, depending on the functions, strength and performance on the spatial database side, desktop GIS and batched processing services. There is some success in a batched processing service enabled by JEQL scripts developed in-house to deal with certain QA tasks that are less efficient by the off-the-shelf software tools.

Population Attributes

After QA work is completed and a final version of polygons is formed, a centroid is generated for each of the polygons. A centroid must be guaranteed inside the intended polygon, *e.g.*, using ST_PointOnSurface, not ST_Centroid in PostGIS; using Centroid (Inner) in Manifold® System.

Through a spatial overlay of the centroids and the update source bedrock polygon table stored in the observation database, the geological unit or unit code is transferred from the polygon table to the centroids stored in the staging database. The geological unit or unit code and other attributes associated with each of the centroids are then transferred to the new and updated polygons.

The final population and QA checking of the attributes for the updated bedrock polygons are carried out by applying the lexicon major and sub-unit templates stored on the database. With a database, this is a process that can be fully automated with SQL scripts.

Consistent map styles and themes are applied to the updated maps and new styles for new geological units or sub-units are added to the metadata table and shared by any users connecting to the same database.

Archiving

Both the newly updated area and the province-wide bedrock maps and contact lines are versioned with time stamps and loaded to the archive database. The maps are consistent and complete within a given mapping area at a time or stage of progression. Data stored in the archive is considered as the BCGS authoritative source.

Production

Products are derived from the authoritative data source maintained by the archive database. Production is carried out by processes designed according to either product specifications or requirements for applications, such as a map at a specific scale or with a specific mineral focus, for visualization or high performance query, etc. The production can be automated with database side triggers or GeoRSS on updates, a topic beyond the scope of this discussion.

CONCLUSIONS

Efficiency in maintaining geological maps can be achieved by leveraging recent advancement in interoperable (and also freely available) spatial database and geospatial tool, and distilling a set of best practices by harvesting available knowledge and expertise in geological mapping.

British Columbia Geological Survey has shown success in prototyping a geology operational database environment (GODE) at a time of limited resources in information technology support and staffing. In-house expertise and the use of free and open source software, including PostgreSQL/PostGIS, JEQL and OpenJUMP, provide a cost effective solution to the GODE prototype.

BCGS is promoting the best practices of map compilation among the staff mapping geologists and some early adoption has helped to refine and expand the list of best practices. Using a phased approach to implement GODE on corporate infrastructure, BCGS builds a fully functional system component into the system architecture before working on the next component. Currently Microsoft® SQL Server® 2008 R2 has been deployed as the back-end corporate spatial database, accessible from desktop GIS Manifold® System and batched data processing engine JEQL. It is already in heavy use for on-going data quality assurance and data integration. The next step is to implement the data quality assurance policies and applications, and to populate the database with required data layers such as the topographic base maps.

ACKNOWLEDGMENTS

Ward Kilby, Larry Jones and Pat Desjardins are thanked for providing insightful comments and carefully editing the manuscript. A portion of this paper is extracted from a technical document on the architecture specifications for the geology operational database environment that was completed by Yao Cui with the support of British Columbia Geological Survey, especially Larry Jones, director of Resource Information Section at BCGS. JoAnne Nelson helped with testing and improving some of the recommended best practices. The development of the system architecture for the geology operational database environment and the compilation of best practices benefited from the QUEST bedrock geology compilation and many discussions with BCGS geologists.

REFERENCES

Brodeur, J. (2011): Geosemantic interoperability and geospatial semantic web; in Handbook of Geographic Information, Kresse, W., Danko, D.M., Editors, *Springer*, in press.

Cooper, A.K., Moellering, H., Delgado, T., Düren, U., Hjelmager, J., Huet, M., Rapant, P., Rajabifard, A., Laurent, D., Iwaniak, A., Abad, P., and Martynenko, A. (2007): An initial model for the computation viewpoint of a spatial data infrastructure; *23rd International Cartographic Conference*, Moscow, Russia.

Farooqui, K., Logrippo, L., and de Meer, J. (1995): The ISO reference model for open distributed processing: an introduction; Computer Networks and ISDN Systems; *Special issue on ISO reference model for open distributed processing*, Volume 27, Issue 8, July 1995.

Hjelmager, J., Delgado, T., Moellering, H., Cooper, A.K., Danko, D., Huet, M., Aalders, H.J.G.L. and Martynenko, A. (2005): Developing a modeling for the Spatial Data Infrastructure; *Proceedings of the 22nd International Cartographic Conference*, A Coruña, Spain.

Hjelmager, J., Moellering, H., Delgado, T., Cooper, A.K., Rajabifard, A., Rapant, P., Danko, D., Huet, M., Laurent, D., Aalders, H.J.G.L., Iwaniak, A., Abad, P., Düren, U. and Martynenko, A. (2008): An initial formal model for Spatial Data Infrastructures; Taylor & Francis, Inc., *International Journal of Geographic Information Science*, Vol 22, No 11, pages 1295-1309.

ISO (1998): ISO/IEC 10746-1:1998-Information technology-Open Distributed Processing-Reference model: Overview.

ISO (2010): ISO/IEC 10746-3 Information Technology-Open Distributed Processing-Reference Model-Architecture.

Logan, J.M. Schiarizza, P., Struik, L.C., Barnett, C., Nelson, J.L., Kowalczyk, P., Ferri, F., Mihalynuk, M.G., Thomas, M.D., Gammon, P., Lett, R., Jackaman W. and Ferbey, T. (2010): Bedrock Geology of the QUEST map area, central British Columbia; *BC Ministry of Energy and Mines*, British Columbia Geological Survey, Geoscience Map 2010-1.

Massey, N.W.D., MacIntyre, D.G., Desjardins, P.J. and Cooney, R.T. (2005): Digital Geology Map of British Columbia: Whole Province; *BC Ministry of Energy and Mines*, British Columbia Geological Survey, GeoFile 2005-1.

Muir T.L., Watkins, T.W. and Berdusco, B.J. (2000): A library of digital line standards; *Ontario Geological Survey*, Open File Report 6026, 208 pages.

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, 58 pages.

OGC (1999): OpenGIS® Simple Features Specification for SQL, Revision 1.1; *Open Geospatial Consortium, Inc.* OpenGIS® Project Document 99-049, Release May 5, 1999, 78 pages.

OGC (2001): OpenGIS® Implementation Specification: Coordinate Transformation Services; *Open Geospatial Consortium, Inc.* OpenGIS® Project Document 01-009, Revision 1.00, Release Date: 12 January 2001, 117 pages.

OGC (2010): OpenGIS® Implementation Standard for Geographic information - Simple feature access - Part 1: Common architecture; *Open Geospatial Consortium Inc.*, OGC 06-103r4, Version: 1.2.1, 93 pages.

RIC (1997): Specifications and guidelines for bedrock mapping in British Columbia; Resources Inventory Committee, *British Columbia Geological Survey*, Information Circular 1997-3, 185 pages.