Multi-Representation Database Development:
Review and Recommendations for National Mapping

Mark Smithgall
mmsmithgall@gmail.com

Abstract

This recommendations paper will review literature and present essential facets one must account for while constructing a multi-representation database (MRDB). Four main aspects must be included in a MRDB: data are stored at different Levels of Detail (LoD), data are capable of being generalized on demand, an update to an object for one LoD shall be conveyed to the other LoDs that portray that same object, and objects must be explicitly linked. Illustrated examples specifically related to hydrographic data are demonstrated with specific reference to literature sources. A hypothetical MRDB interface is created, recommendations are made for national mapping agencies (NMAs) and suggestions for future work are made.

Keywords: MRDB, multi-representation database, multi-resolution database, multi-scale mapping, feature representation, data modeling, level of detail (LoD), linking

1. Introduction

The multi-representation database (MRDB) is the new answer to the question of how to store, organize and serve large amounts of geospatial data. MRDBs are used to "store the same real-world phenomena [for example, a bridge, river, road, or anything else portrayed on the map] at different levels of precision, accuracy and resolution" (Hampe et al., 2003, 1). In addition, MRDBs are able to store phenomena so that they are capable of being presented and symbolized in different manners on the fly.

Level of Detail (LoD) data are generalized for one theme (such as hydrography) to suit a particular range of display scales within the context of a database that includes multiple LoDs for multiple themes. A round-number map scale that suggests the large-scale end of the range for suitable display scales is referred to as the *base resolution*. Three data-frame screenshots are displayed in Figure 1 for the three base resolutions of the National Hydrography Dataset (NHD): 1:5,000 (Local), 1:24,000 (High) and 1:100,000 (Medium). Additional LoDs are developed from these existing base resolutions. For example, to create an LoD at 1:50,000, the NHD High resolution dataset (i.e., one of the three *base resolutions*) would need to be generalized to 1:50,000 scale.

MRDBs have four distinct advantages over regular GIS databases and other traditional means of data storage and manipulation: (1) data are stored at different levels of detail, (2) data are capable of being generalized automatically on demand, (3) an update to an object at one LoD shall be conveyed to other LoDs that portray the same object at their respective base resolutions, and (4) objects must be explicitly linked to one another so the MRDB recognizes its corresponding objects at other LoDs.
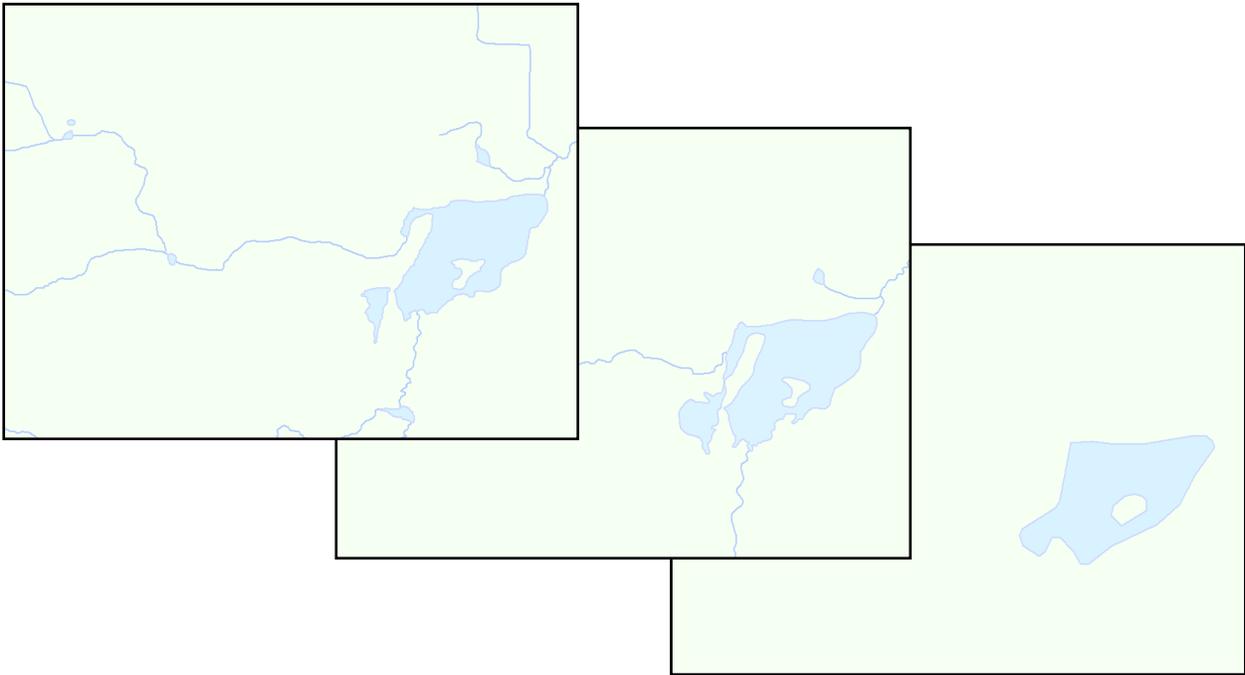
Figure 1: NHD Local, Medium and High resolution data serving as three base resolutions.

The potential advantages of MRDBs are many: decreased "load-time" for wireless users; immediate updates for data revisions, whereby a change to a base LoD propagates automatically throughout the rest of the database; and decreases in processing time for data modification because generalizations of a series of linked LoDs are conducted automatically rather than generalizing every LoD, one at a time as a separate database for each resolution. In addition, MRDBs provide a data clearinghouse enabling an organization to centralize and streamline their data management procedures.

This report will concentrate on using literature sources to advise the creation of an MRDB. The underlying concepts behind MRDBs will be explained and suggestions will be offered focusing on different aspects of MRDB design and construction.

2. Review of Literature

The focus of the project has been on finding, assessing, and compiling relevant literature sources to aid in envisioning MRDB creation. I did not find journal articles, books, case studies or proceedings papers with instructions on how MRDBs are created. One paper documenting the use of a Federated Database Management System (FDBS) regarding MRDB application was found (Anders and Bobrich, 2004); however, I did not find literature documenting the use of ESRI or Microsoft software (common in industry and government mapping) for accomplishing this task. A plan of action could be to use the literature as guidance to begin to plan data characteristics that allow for later MRDB formats and encourage creation of MRDB tools.

Re-generalisation and construction for automated incremental updating

The proceedings paper by Bobzien et al. (2005) was presented at the *Twenty-second International Cartographic Conference* in Spain. The authors present a case study with precise attention given to buildings portrayed on topographic maps and to National Mapping Agency (NMA) concerns. Manual editing, which is extremely expensive in money and time, needs to be minimized in order for MRDBs to function automatically. The problem exists that no computer or algorithm can efficiently and effectively take the place of human oversight at this time.

The authors describe the challenge of automated generalization. They present the problem of incremental updating with MRDBs: "Updating, performed on the dataset of the higher resolution, needs to be propagated automatically to the datasets of lower resolutions" (Bobzien et al., 2005, 1). If one edits or adds a building to the highest resolution LoD, how would those changes be reflected in the other *explicitly linked* LoDs? Explicit linking occurs when each feature in each LoD is linked so the DBMS (database management system) can communicate updates and feature representation calls between different LoDs at different resolutions. For example, in Figure 2 the high-resolution LoD buildings are separate entities, but in the medium-resolution LoD (at right) the buildings are amalgamated.
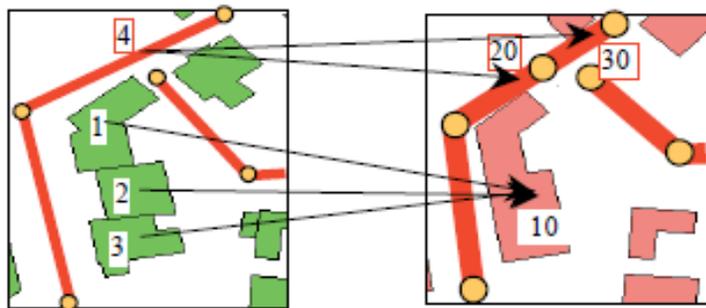

Figure 2: Explicit linking (Hampe et al., 2003, 5, authors' Figure 6).

The final concept introduced by the authors is how time affects the update process. Figure 3 shows how the *time dimension* and updates (x-axis) relate to the *resolution dimension* and generalization (y-axis) throughout the generalization process. Two approaches are given: re-generalization which "reapplies the initial generalisation to the updated feature," and construction which "reapplies the update to the generalized feature" (Bobzien et al., 2005, 6). The authors then conclude the proceedings paper with a summary of their case study results: presenting a *novel generalization relation class* in the features' attribute table created to store the history of any generalization operators that were executed on the feature.
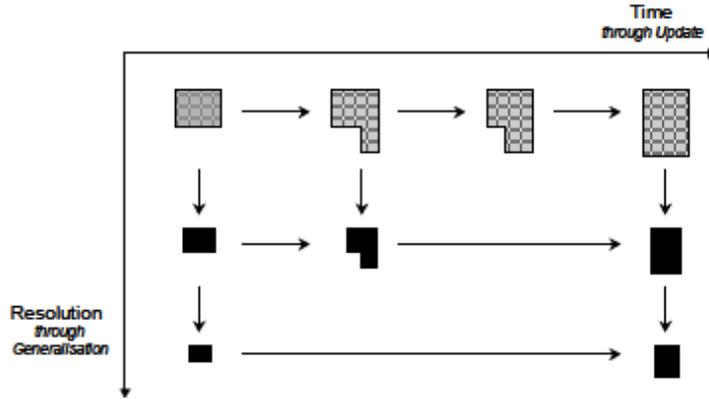
Figure 3: The time dimension related to the resolution dimension (Bobzien et al., 2005, authors' Figure 5).

Generalization for on-demand web mapping by multiscale databases

At the *Tenth International Symposium on Geospatial Theory, Processing and Applications*, Cecconi et al. (2002) describe a concept related to modern, albeit theoretical, MRDBs – the generation of maps on demand. Such capability will be an advantage of an MRDB once implemented. The authors also describe the use of a multi-scale database and cartographic generalization methods.

As the oldest article drawn upon for this report – published in 2002 – it is dated in the technology available at its press time. The article describes the precursor to the development of a modern MRDB: a Multi-Scale Database (MSDB). They also describe a concept vital to use of an eventual MRDB: explicit linking between different LoDs. The defining principle behind a MSDB (and eventually an MRDB) is that "every object in one scale has to know its counterpart in the larger or smaller scale" (Cecconi et al., 2002, 4).

The links between objects can then be used to classify any generalization operators to be used or to preserve the topological relationships between objects, LoDs, or the actual links themselves. Figure 4 shows two base scale LoDs containing original, digitized data: one at 1:200,000 scale and the other at 1:25,000. The third LoD at 1:100,000 is created from the explicit links to the two base-scale levels. Mustière and Smaalen (2007) term the created LoD the *composite LoD*.
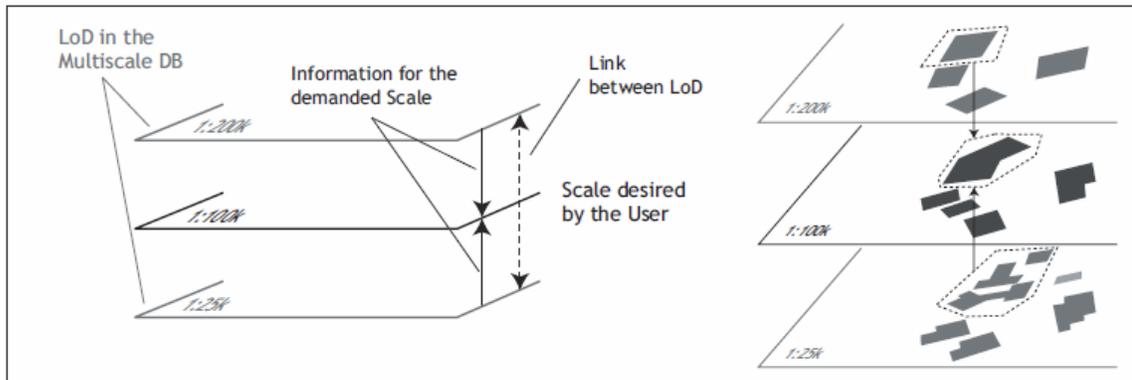

Figure 4: Explicit linking between different LoDs (Cecconi et al., 2002, 5, authors' Figure 1).

Classifications of generalization operators by scale range are used to design the contents of the MSDB itself. "In particular, aspects like 1) the levels of detail, 2) the limits of applicability for the LoDs of the MSDB, and 3) the generalisation operators used to transform between the LoDs can be addressed" (Cecconi et al., 2002, 6).

MRDB applications for data revision and real-time generalization

From the *Twenty-First International Cartographic Conference* in South Africa, the proceedings paper by Hampe et al. (2003) provides an excellent abbreviated definition of MRDB systems and what their capabilities should encompass. "Different views of the same physical objects or phenomena can be stored and linked [in a MRDB]… stemming from different views of the world, different applications, as well as different resolutions" (Hampe et al., 2003, 1).

Two main aspects are described: the storage and explicit linking of data in the form of LoDs. The second aspect is the defining concept of MRDBs, *real-world objects* must *be explicitly linked to each other through the different LoDs in order for each object to "know" its complimentary objects at other LoDs* (Figure 5). Furthermore, the authors suggest two different cases of using data for designing a new MRDB: whether to "link existing data sets of different scale or thematic concept by matching procedures; or to create new data sets from existing ones [data sets] then form new layers in the MRDB" (Hampe et al., 2003, 2). Linking existing datasets should prove to be less expensive; however, any inaccuracies present in the data will be conveyed to the different LoDs in the MDRB. Creating new datasets would increase the expense; however, this case could allow the user to verify the accuracy of their data.
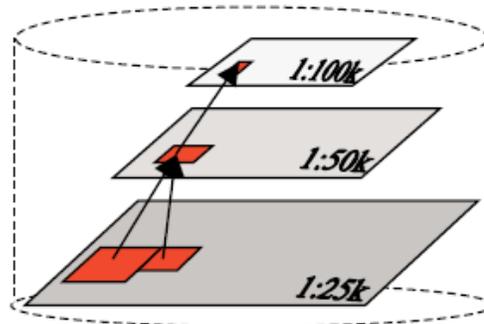

Figure 5: Explicit linking (Hampe et al., 2003, 2, authors' Figure 1)

The authors also present two case studies concerning MRDBs: The Project WIPKA and The European Union Project GiMoDig. Project WIPKA, developed by the German Federal Agency of Cartography and Geodesy, is concerned primarily with associating topography with MRDBs and developing concepts for updating LoDs. Project GiMoDig (Geospatial Info-mobility Service by Real-time Data-Integration and Generalisation) intends to improve data delivery for NMAs in database form. Finally, the authors present ideas on how the data update process should be accomplished within an MRDB.

Database requirements for generalization and multiple representations

The chapter by Mustière and Smaalen (2007) comes from *Generalisation of Geographic Information: Cartographic Modeling and Applications*, a reference book edited and sponsored by the International Cartographic Association. The chapter provides an overview of data modeling

requirements and the general concepts of databases. It also details the pros and cons of MRDBs and their application to the requirements for generalization throughout the MRDB process of duplicating real-world elements at varying scales.

An introduction acquaints the reader with the basics of database management: how databases save, query, present, and transform the chosen data and how they hold shape, location, attribute information, and spatial relationships among data. The authors chronicle the advantages of MRDBs: decreased "load-time," instantaneous updates, the ability to greatly decrease processing time for data modification, and serving as a central data clearinghouse. The last major concept offered by Mustière and Smaalen (2007) is their summary of an ideal generalization database system: a sort of "wish-list" of features that do not all concurrently exist in a current database system:

> Support the use of "rich conceptual models" of geographic data to support complex data modeling and processing, rely on database management system (DBMS) to take full advantage of the data management capabilities, support use of classification hierarchies and inheritance to better organise data and better guide generalisation processes, allow the building and retention of links between component and composite objects, support topology that allows the relationships of objects to be queried, and deal with multiple representations for a single real-world entity (p. 135).

Their recommended data schema concentrate more on natural organization, rather than management of geometries. This simplifies the data storage process and enables further data manipulation, like generalization actions. The main emphasis of Mustière and Smaalen (2007), however, is that a DBMS must be used to take full advantage of the beneficial data management capabilities such as versioning, querying, and optimization. Additionally, classification hierarchies (such as stream order for a hydrologic network) must be created and maintained to provide the desired data management capabilities.

3. The MRDB Concept

Four main principles established in the literature review that distinguish an MRDB are summarized with reference to NHD examples: data storage, generalization on demand, automatic updates, and explicit linking. The topics of incremental updating and alternative feature representations are also elaborated with NHD examples.

MRDB Principles

Data Storage:
    A main purpose for any size and type of database is to provide a central location to store data. With an MRDB, data are stored with different LoDs. Figure 6 illustrates an NHD example in which features stored in different LoDs are linked and shows how a composite LoD would be made. Figure 6a is the NHD Medium data intended for 1:100,000 mapping without any further generalization or modification (raw); Figure 6b is the raw NHD High data intended for 1:24,000; Figure 6c is a composite LoD, possibly 1:50,000, from the generalization of Figure 6b.
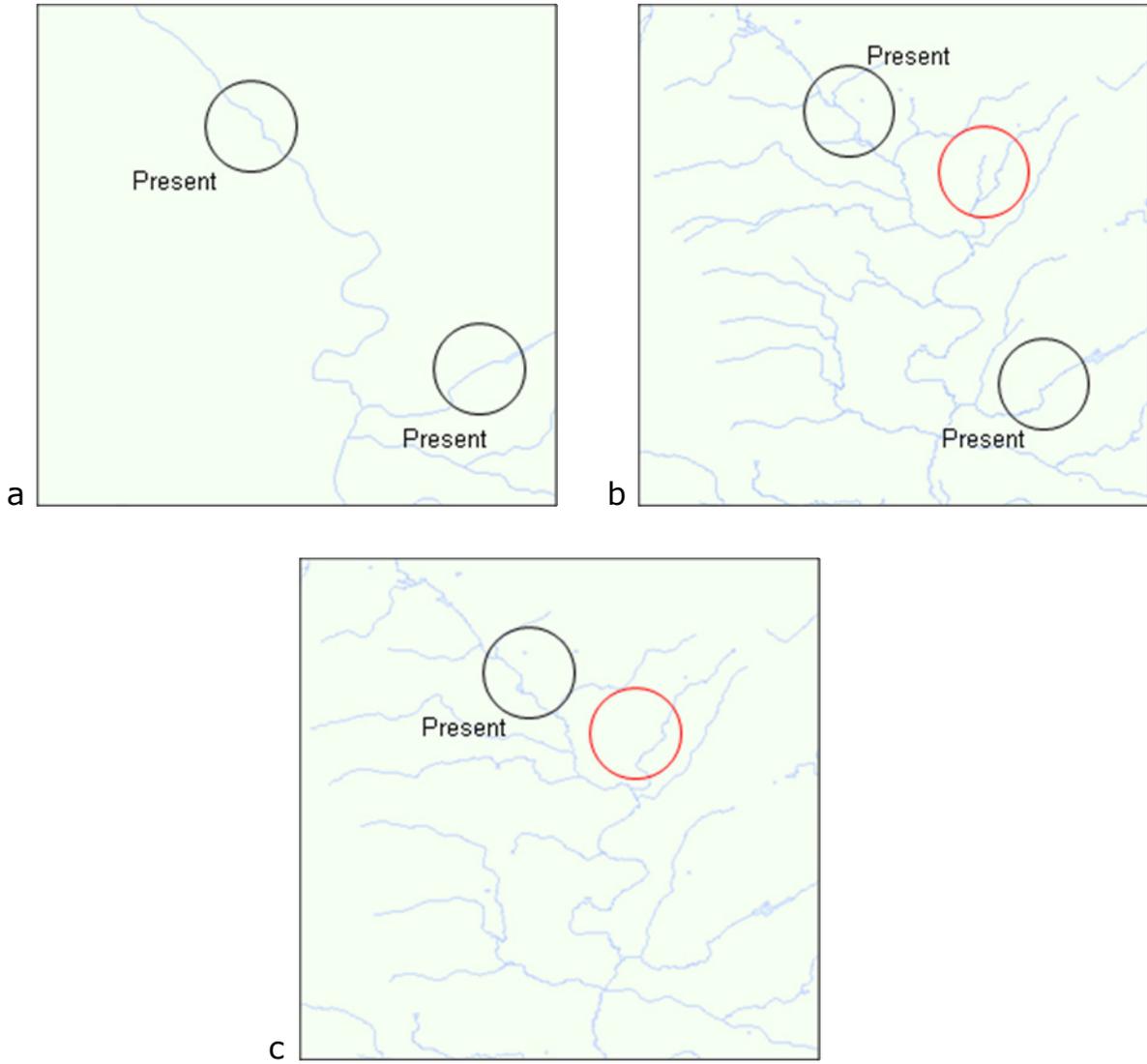
Figure 6: National Hydrography Dataset linking example at three LoDs: a) NHD Medium, b) NHD High and c) the composite ≈1:50,000 scale LoD.

Note the features that are present in all three LoDs represented by black circles and noted as "Present." An example feature that is not conveyed through LoD processing due to modification/generalization is notated with a red circle.

Generalization on Demand:
MRDBs must have the ability to be generalized on demand to suit a multitude of situations. Whether the situation dictates the data be portrayed at a small scale, possibly 1:1,000,000 or the data are represented at a scale possibly as large as 1:2,500 – the base LoDs must have the ability to be generalized automatically and to work with the composite LoDs to suit these varying needs.

Figure 7: Comparison of generalized NHD to non-generalized NHD. This could be viewed as the scale change from 1:200,000 (left) to 1:50,000 (right)

Figure 7 portrays a small area of NHD High data. The data within the red box (the left third) is shown after manipulation by a simple line smoothing algorithm. The remaining two-thirds are ungeneralized data. If detailed LoDs inside MRDBs do not lend themselves to on-the-fly generalization, this approach could be applied to on-demand zoom levels.

Automatic Updates:

An update to a feature in a base LoD (Mustière and Smaalen, 2007) shall be conveyed to the other composite LoDs which portray that feature. If a building's geometry was modified (refer to Figure 3) that update shall automatically be conveyed through to the subsequent LoDs of coarser resolution.

Explicit Linking:

Within the MRDB, objects (features such as buildings, roads, and streams) *must* be explicitly linked to corresponding objects at other LoDs. This is the central theme for an MRDB and, with existing technology, it is hard to accomplish. A feature will be stored in the MRDB and the DBMS will understand which LoD will be used to portray the feature based upon the display scale desired by the user because the objects that represent it are linked (refer to Figure 8 for a visual example). The low-resolution LoD could contain an object for a suburb of a city, the medium LoD could offer city blocks, while the high LoD represents individual buildings. Explicit links are built into the MRDB so that when changing LoDs, the DBMS knows which features correspond to each object.
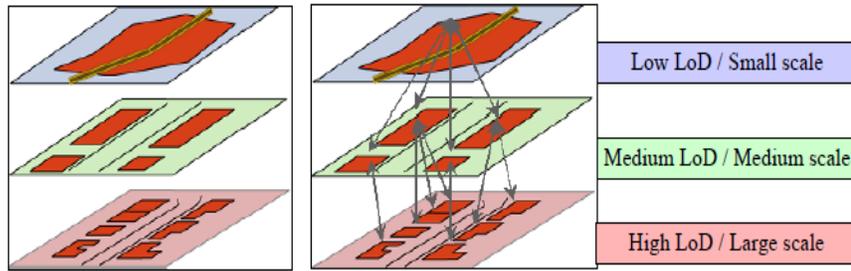
Figure 8: Explicit linking of city data (Hampe & Sester 2004, 1, authors' Figure 1).

In Figure 9 (a repeated presentation of Figure 2 seen earlier), two types of objects are portrayed: buildings and roads. Building 10 in the medium LoD is stored as three entities (buildings 1, 2 and 3) at the high LoD. Conversely, road segment 4 at the high LoD is broken into road segments 20 and 30 at the medium LoD. When changing viewing scale, the DMBS understands which feature(s) to call, from the high LoD or medium LoD.
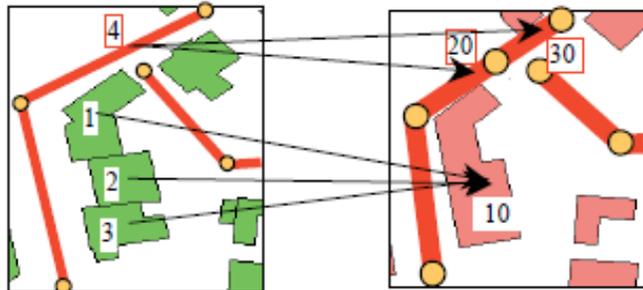

Figure 9: Explicit linking of road and city data, high LoD on left, medium LoD on right (Hampe, et al., 2003, 5, authors' Figure 6).

Figure 10 portrays the explicit linking concept specifically with regard to NHD data before further generalization. Lake Shaftsbury is portrayed in all three LoDs (NHD Local, High and Medium). The red arrows visually depict the link between features. The lake is shown in all three; however, the river draining from the south is only shown in the Local and High LoDs. The small lake, noted with the red triangle, is only shown at the Local LoD. For actual MRDB implementation, linking based upon reach code could be a practical and usable method of applying this concept, enabling the network topology to be maintained.
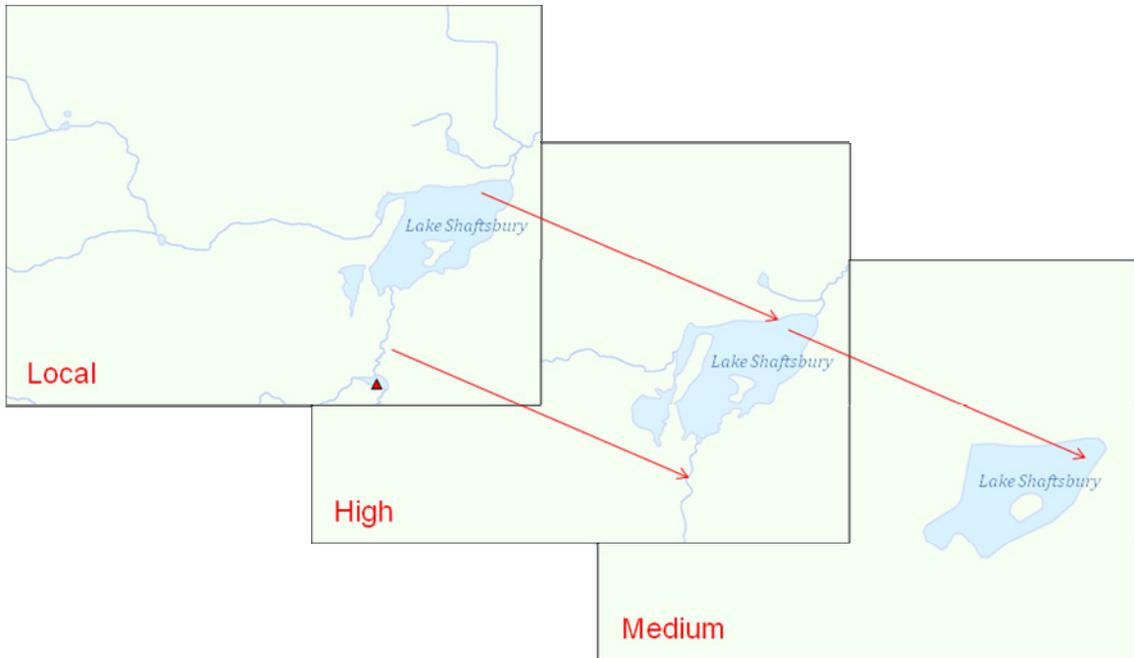
Figure 10: Explicit Linking (NHD example showing waterbodies and streams).

Incremental Updating

Bobzien et al. (2005) describe a concept vital to the proper function of MRDBs: incremental updating. Two ways of approaching this process exist: re-generalization and construction. In Figure 11, three operations are taking place: enlargement (E), simplification (S) and updating (U). On the left (11a), the ***re-generalization*** approach is taken: the original building is portrayed
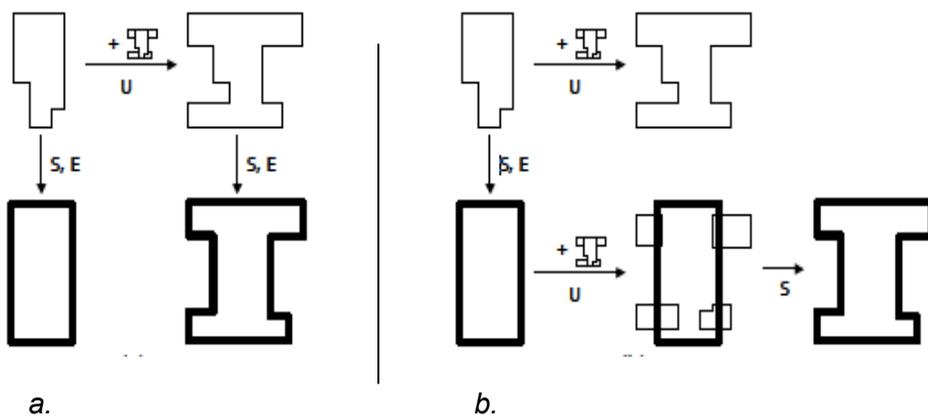

a.   b.

Figure 11: Two types of incremental updating: re-generalization (a) and construction (b) (Bobzien et al., 2005, 6, authors' Figure 8).

in the upper left. Its original generalization is located directly beneath it on the lower left. When an update is needed, the building additions are added to the original feature and then the entire feature is re-generalized. The entire generalization is re-done.

On the right hand side of the figure (Figure 11b), the *construction* process is carried out. Instead of adding building additions to the original building, these additions are added to the already generalized building – bottom row, in the middle. The result of the addition is then generalized; the final product is shown on the bottom right.

NHD example of Re-Generalization:

When an update requires re-generalization, the DMBS applies the update to the original ungeneralized feature. In Figure 12, the original NHD data (12a) is supplemented with detailed first-order streams (12b). This "stubble" is not generalized (for the purposes of illustration, these first-order streams were hand drawn randomly and are not intended to accurately represent existing streams). After the ungeneralized "stubble" is added to the original data, the entire dataset is then generalized (12c).

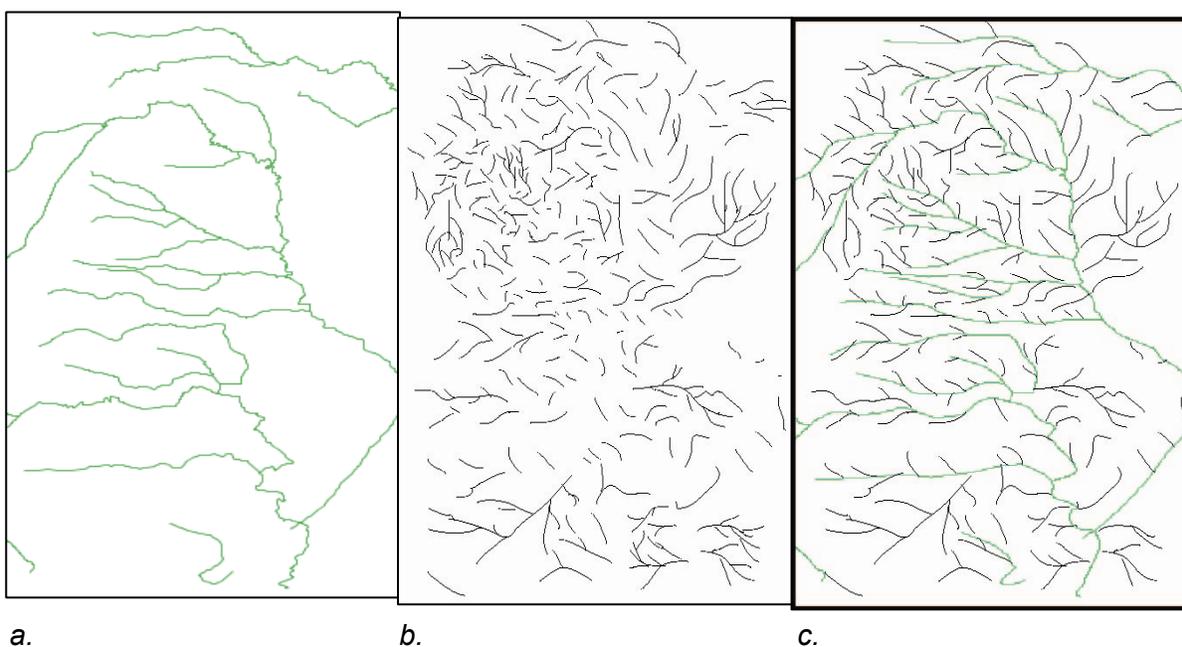a.   b.   c.

Figure 12: Re-Generalization (incremental updating) of NHD flowline example. Original NHD data (a), detailed first-order "streams" (b), and result with all features generalized together (c).

NHD example of Construction:

Alternatively, stream updates may be constructed by adding them to previously generalized features. This process applies the initial generalization operators to the updated features only. In Figure 13, an already generalized version of the NHD is the starting data (13a). Next, the "stubble" is added (13b); however in this construction case, the "stubble" is pre-generalized. The final result need not be further generalized (13c).
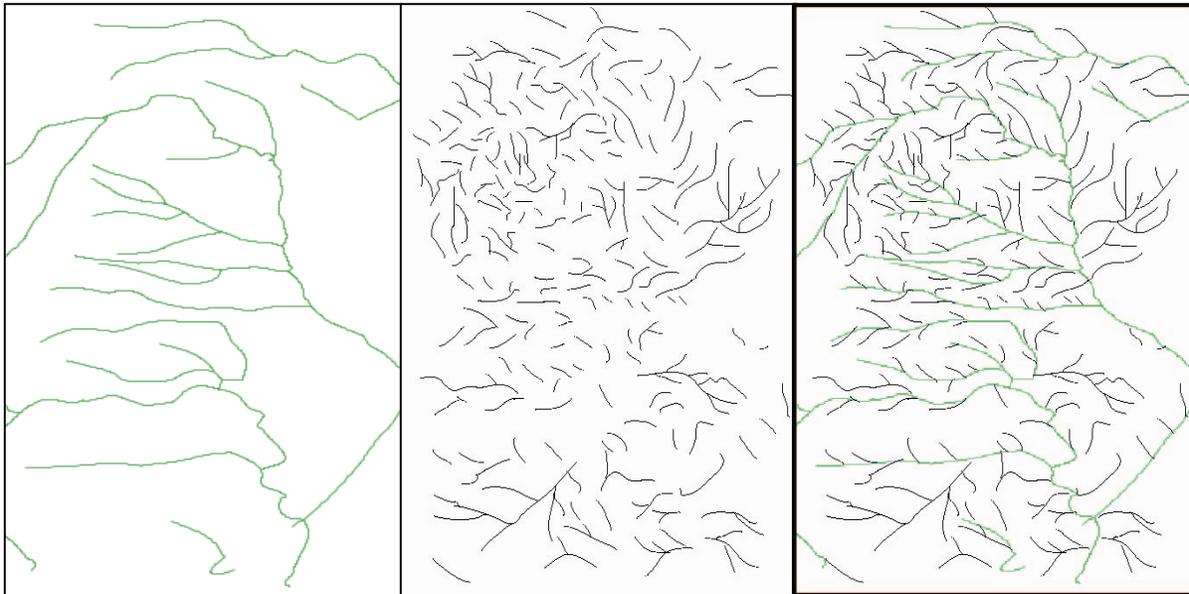
a.   b.   c.

Figure 13: Construction (incremental updating), NHD flowline example. Generalized NHD data (a), generalized first-order "streams" (b), and result with generalized features together (c).

Alternative Feature Representations

With the future development of MRDBs, an efficient way to correctly represent features from the database may be created. As one zooms through the LoDs, different features and feature groups will be turned on and off, they will change symbology (colors, geometry, size, etc.) and will be represented in different fashions to suit the scale of display. This flexibility enables creation of few LoDs, rather than an LoD for every map scale level. Additionally, these feature representations will have the ability to be tailored to the purpose of each map. If one wanted a topographic map highlighting the location of Arches National Park, as opposed to a highway map of Utah, gas stations would not be portrayed. On a 1:1,000,000 scale map, only major rivers will be portrayed and first order streams will not be shown. On a 1:10,000 scale map, first order streams will be shown; however, the major rivers will be so large at this scale they change from line segments to polygons reflecting this change.

In Figure 14 a scenario is given in which three different base scales will possibly be portrayed using GIS tools: 1:100,000, 1:24,000 and 1:5,000 (NHD Medium, High and Local resolutions, respectively). For example, as one zooms through these LoDs, the Medium LoD "Main River" layer splits into the High "Main River – Downstream" and "Main River – Midstream." The High "Main River – Upstream" layer splits into Local "Main River – Upstream" again as well as Local LoD "Small [water] Falls." This is the type of functionality a GIS system that uses an MRDB could offer users doing multiscale work.
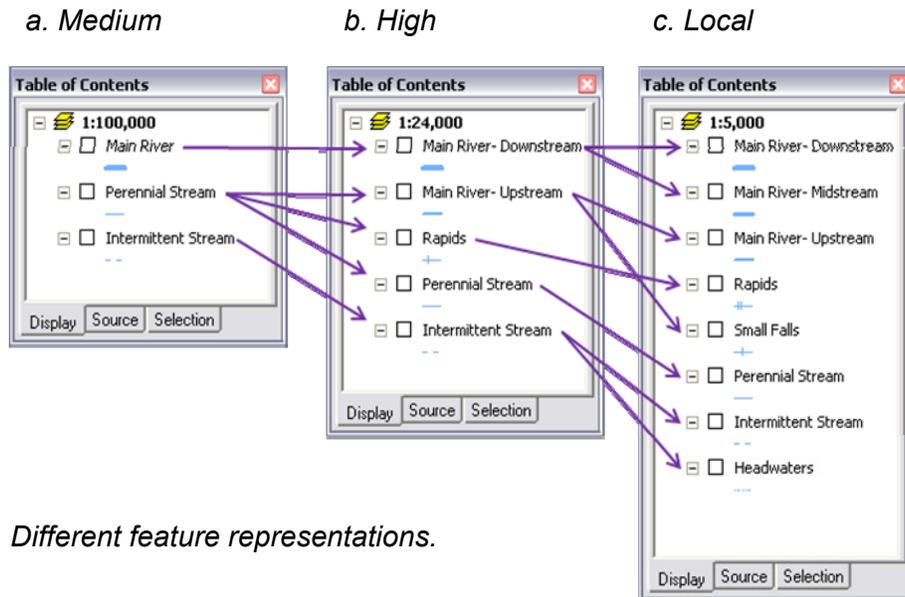
a. Medium   b. High   c. Local

Figure 14: Different feature representations.

4. Recommendations in National Map and NHD Context

The benefits of implementing a multi-resolution database by a national mapping agency to store and serve data could outweigh the expense to create and maintain one. As more datasets of varying complexity and detail become widely available, problems will be encountered with database speed and efficiency serving data to customers.

I presented research completed by individuals and organizations, some in the context of NMAs; however, the true use of MRDBs is, for the time being, just a dream. A collaboration between NMAs, database companies and GIS-software companies with the intent to produce an MRDB should make substantial progress.

The USGS goal in 2015 is to have 7-day turnaround time on data updates for the National Map. The use of MRDBs to accomplish such fine temporal resolution will be key. NHD data are organized based upon the widely used reach code, so the addition of a linkage column in the data's attribute table to explicitly link features from different LoDs is the next step forward in MRDB development for USGS. Hampe et al. (2003) discuss Project GiMoDig: the attempted creation of an MRDB for topographic map use for European Union mapping agencies.

Currently, the most challenging aspect of MRDBs is the inability to explicitly link and relate features. Once this task is accomplished, the implementation of MRDBs could become prevalent.

5. Conclusion and Future Work

The basis of the project thus far has been on finding, assessing and compiling relevant literature sources to aid in envisioning MRDB creation. A set of guidelines has been laid out to serve as the "wish-list" for ideal MRDB features: (1) data are stored at different LoDs; (2) data are capable of being generalized automatically on demand; (3) an update to an feature at one LoD shall be conveyed to other LoDs that portray the same object at their respective base resolutions; and (4) objects must be explicitly linked to one another so the MRDB recognizes its corresponding objects among all LoDs.

MRDBs are an appealing solution to many types of geospatial data storage and service problems. Presently, MRDBs are a theoretical solution. The software needed to link features between different LoDs does not yet exist using commercially available GIS software. I anticipate further research stimulating the development of MRDBs will be carried out and the development of an MRDB is imminent.

Resources

Anders, Karl-Heinrich, & Joachim Bobrich (2004). MRDB approach for automatic incremental update. Proceedings of ICA Workshop on Generalisation and Multiple Representation. Leicester, United Kingdom, August 20-21.

Bobzien, Matthias, Dirk Burghardt, & Ingo Petzold (2005). Re-generalisation and construction – two alternative approaches to automated incremental updating in MRDB. Proceedings of XXII International Cartographic Conference, A Coruña, Spain, July 9-16.

Brewer, Cynthia, & Barbara Buttenfield (2007). Framing guidelines for multi-scale map design using databases at multiple resolutions. *Cartography and Geographic Information Science*, 34(1): 3-15.

Cecconi, Alessandro (2003). *Integration of cartographic generalization and multi-scale databases for enhanced web-mapping* (Doctoral dissertation). University of Zurich, Switzerland.

Cecconi, Alessandro, Robert Weibel & Mathieu Barrault (2002). Improving automated generalization for on-demand web mapping by multiscale databases. Proceedings of 10th International Symposium on Geospatial Theory, Processing and Applications, Ottawa, Canada.

Chaudhry, Omair, & William Mackaness (2006). Modelling geographic phenomena at multiple levels of detail. Proceedings of AutoCarto 2006 Research Symposium. Vancouver, Washington, United States, June.

Hampe, Mark, & Monika Sester (2004). Generating and using a multi-representation database (MRDB) for mobile applications. Proceedings of ICA Workshop on Generalisation and Multiple representation. Leicester, United Kingdom, August 20-21.

Hampe, Mark, Karl-Heinrich Anders & Monika Sester (2003). MRDB applications for data revision and real-time generalisation. Proceedings of XXI International Cartographic Conference, Durban, South Africa, August 10-16.

Mustière, Sébastien, & John van Smaalen (2007). Chapter 6: Database requirements for generalisation and multiple representations. In Mackaness, W., A. Ruas, and T.L. Sarjakoski (Eds.) *Generalisation of Geographic Information: Cartographic Modeling and Applications* (pp. 113-136). London: Elsevier.

Petzold, Ingo, Dirk Burghardt, & Matthias Bobzien (2006). Workflow management and generalisation Services. In 9th ICA Workshop on Generalization and Multiple Representation, Portland, Oregon, United States.

Thomsen, Andreas, & Martin Breunig (2007). Some remarks to topological abstraction in multi representation databases. Proceedings of International Workshop on Information Fusion and Geographical Information Systems, IF&GIS 2007, St. Petersburg, Russia.

———

Paper revised, edited and reformatted by Cynthia Brewer