A Product Driven Approach to Designing a Multi-Purpose Multi-Scale GIS Base Map Database that Supports High Quality Mapping

Charlie Frye
380 New York Street
Redlands, California 92373
cfrye@esri.com
v. (909) 793-2853
f. (909) 307-3067

A Product Driven Approach to Designing a Multi-Purpose Multi-Scale GIS Base Map Database that Supports High Quality Mapping

Abstract

The term "base map" has many specific meanings to a variety of organizations, products, and processes, but in the context of organizations who use GIS and who produce and publish maps, a base map contains the data that underpins all the map products and their common workflows in the organization. This paper describes a management strategy for the elements of a GIS base map critical to ensuring that an organization is successful in their map making work. A key to a successful multi-purpose multi-scale base map is a product driven approach. To illustrate this approach a hypothetical example of a local level governmental agency is used to show how the maps an organization will produce are used to inform the database design process. This example is based on real world experience working with many such organizations. This paper also contains discussions of several topics that impact multi-scale, multi-purpose base map data model design such as implications for data capture and deriving data for smaller scale maps. These ensuing discussions provide a logical experience-driven basis for making better strategic decisions about base map data modeling.

Introduction

A base map means many things to many people, but in the context of GIS and mapping, most agree that it contains the data that underpins many of the maps a given organization produces. From a resource management perspective, a base map should support efficient workflows within an organization as they seek to produce their products, and therefore, the "goodness" of a base map is indicated by the number of products using its data. Thus, good base maps are appropriately informationally rich and current. "Appropriately," here, means the standard for data richness and currency are based on client expectations for the products a particular base map will support.

The focus of this paper is on base maps that a typical municipal or county government would create, maintain, and use to make maps that present the state of their area of responsibility to the public and to decision makers. Such maps include traditional paper maps and interactive maps that may be purely informational or the basis for other services. The data in these base maps are assumed to be collected, verified, and managed locally. Ideally, data from many local base maps are combined to create state, provincial, and national base maps. Thus, good base maps can be used to provide needed information to decision makers at all levels of government. For example, data to support emergency response efforts for events over large extents, such as natural disasters including earthquakes, tsunamis and hurricanes, would ideally come from these detailed local base maps. Disasters that occur over smaller extents, such as landslides, some floods, and some fires, also require the high resolution information in local base maps. This need exists now and will increase in the future because increasing population pressure is at the root of increases in expectations for faster and better emergency response and management.

Up-to-date, accurate maps and their associated base maps are essential components in successful emergency response efforts because maps with current and detailed information allow resources to be rapidly deployed and effectively managed. Such base maps have certain essential characteristics; the following are hallmarks of a good base map:

1. Rich and appropriate content exists for all of an organization's designed standard needs and products. Another way to say this is that the products and services an organization provides drive the content requirements for their base map.
2. Data should be correctly spatially integrated. There are several kinds of topological relationships (such as those that prevent features from overlapping or intersecting, and those that maintain connectivity between other features) for much of the world that we should ideally capture in a base map. For instance, street centerlines should never cross parcel lines—in fact they should represent, as closely as possible, the middle of streets; street centerlines should connect when the streets they represent intersect; buildings should not overlap roads; a river's gauging station should be inside the river polygon; etc.
3. Information should be modeled first to suit the purpose and requirements of all the organization's standard maps and analytical products, and secondly, if possible, to support efficient management of the base map and products. To illustrate this idea, consider that in the United States in 2003, Executive Order Hspd-7 (Bush 2003) on Critical Infrastructure Identification, Prioritization and Protection was issued. The location of these places or facilities is often very specific—typically more specific than the detail that most communities already manage in their parcels, buildings, or address data. For example, a chemical plant may have several facilities of critical nature within its plant boundaries and parcels, but buildings and addresses are not sufficiently spatially referenced to locate such places. Rather than making a tabular reference in those existing datasets which would result in specific locations not actually being captured or managed, new data should be spatially integrated and should include attribute relationships as needed.
4. A consistent, straightforward, data attribution scheme should be used throughout the base map database. Attribute names should consistently be used for the same purpose throughout the database.
5. A base map provides a centralized repository for information management. Data such as place names, lineage, history, ownership, and contacts are ideally managed once and only linked as needed by individual products to other data in the base map.

Base maps with these hallmarks are not developed organically or accidentally. Instead a broadly considered and informed information management strategy is used. This strategy is product- and therefore mission-driven with the intent of streamlining production and maximizing the effectiveness of the organization's resources. This rest of this paper lays out the basic strategy for good base map design and discusses issues that impact GIS base maps that support cartography and mapping. The method described is derived from experience working with several local governmental agencies that are responsible for building base maps and making cartographic products. Experience is also drawn from working with higher level governmental agencies and departments that are responsible for compiling both spatially continuous base maps for larger areas, as well as smaller scale base maps from local base map data.

Method for Building High Quality Base Maps

The way an organization creates its base map has a profound impact on the quality and utility of any publicly disseminated information that will come from that base map. All people in the organization who are responsible for the collection and management of data that will be used in products and stored in the base map should collaborate on the requirements that underpin each product, opportunities to share responsibilities, and prioritization of their products and the workflows. This collaboration is the basis for understanding the organization's requirements in a non-redundant fashion. While the implied target of such planning is the portion of an organization responsible for a base map, this strategy can also be scaled and applied to subsets of organizations with a large staff and many departments.

In the beginning, this is ideally a conversation or planning meeting involving all parties that will share a common base map to discuss the requirements for each of their products. Also, it is often useful to have an outsider, organizationally speaking, participate in this meeting. This should be somebody who knows the organization's goals but is not part of the day-to-day politics of the organization. The purpose for this is to provide a respected non-threatening party with whom everyone will communicate a full explanation or justification of each requirement, and effectively minimize errors of assumed communication. This sort of collaborative or 'what can we learn from each other' planning was illustrated in Brand (2001) as one logical outcome of the September 11, 2001, terrorist attacks in which the relative stability of the New York City's electrical network was evaluated and contrasted with other systems, particularly telecommunications, which fared worse. Mapping and data production for mapping can require significant investments in staff and related resources, which, in the public sector comes at fiscal and political cost, thus making informed alliances between groups of staff a valuable commodity.

The following paragraphs describe this sort of planning process using the simplified context of a county governmental agency that needs to make three map products. In practice, this process starts with a list of all the products, including paper and interactive maps, as well as analytical functions that are map based. In this example product X is a 1:2,500 scale ortho image based map. The map has lines that represent various boundaries and text that labels buildings, boundaries, and other cultural features. The map is used by the public safety agencies within the county. This map also includes points showing critical infrastructure locations. Product Y is a 1:5,000 scale reference map. The map has a contour line base shown over a hypsometrically tinted terrain model. Roads, hydrography, boundaries, vegetation, and cultural features are shown and labeled. One specific requirement is that it shows block by block street address ranges. The map is the basis for planning activity within the county and serves as the cartographic base for many public and private maps for the county. Product Z is a 1:25,000 scale topographic map. Within the county there are scenic parks that include some rugged terrain and many trails. This has spawned an avid orienteering community and the county produces a set of USGS-style topographic maps to facilitate them, and when needed, their rescuers.

Generally, this list should include those products or map based services that are consistently or regularly produced; one-off or special projects should not be included. Initially, it is also useful to narrow down this list to only the most important products the organization produces—this methodology can be scaled to accommodate new products when and as they are needed.

Basic Organization of Data or Information Themes

The base map planning process must have a sound basis for discussing data. Several years of data modeling research at ESRI using standard data modeling techniques has yielded several common threads. One of these is a common thematic organization of data (Arctur and Zeiler, 2004). Conceptually speaking, all base maps contain data from themes, like transportation, hydrography, and administrative boundaries. Figure 1 shows an example geodatabase that contains feature datasets used to store the data for each of these themes. These themes are used throughout this paper.

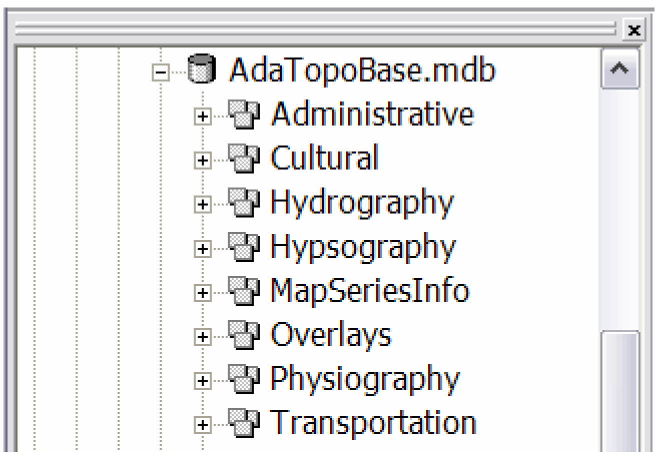

Figure 1. Example geodatabase with thematic data organization. AdaTopoBase.mdb is personal geodatabase and each theme (e.g., Cultural, Hydrography) is a feature dataset. Specific feature classes are stored within these feature datasets.

Figure 1 illustrates a sensible way to organize base map data because it allows information that typically requires spatial integration to be managed together. In a geodatabase a feature dataset can also store a topology or a geometric network to manage the spatial integrity of that data. The naming convention is easy to understand; casual users of this database can easily browse and find data of the sort they are looking for or quickly know that these data are not in this database. Too often this top level of the database design is omitted resulting in most feature classes being stored at the root level, and feature datasets being created only for the spatial management function they fulfill rather than the kind of information they contain. As more people in an organization come to work with its data, the cost of training and explaining the data organization increases dramatically without this kind of easy to follow data naming and organization.

There are also cases where products, particularly maps that may require inter-theme spatial integration such as topological relationships between streets and buildings, would obviate some behind-the-scenes organization that differs from what is shown in Figure 1. When possible, these cases can be handled as part of the product production workflow; for example, while a topology could certainly be used to find all the buildings that overlapped road centerlines, a selection by location function or a graphical conflict detection function applied at the appropriate time in the workflow would have a much lower impact on the rest of the organization using the base map.

Product Data Requirements

The next step in the base map planning process is to enumerate the kinds of data that each product requires. Figure 2 contains an inventory of data that are organized using the example products and themes that have been discussed thus far. This inventory is a top level description of the data requirements for the three map products. To create this type of list, the products are inventoried to determine which data themes are needed. The basic idea is that each product, to fulfill its purpose, requires these elements of contextual information from the base map.

Beginning with each product's needs, this examination of the data content for specific products can be based on a matrix that delineates the specific GIS feature classes needed on a product-by-product basis. An example matrix is shown in Figure 2. Along the X axis, columns were created for all the products and maps. Along the Y axis, a row was added for each feature class (type of data and type of geometry) required for the product in the first column. Check marks indicate data that will be used for each product. Rows were added when another feature class was required and columns were added when another product was added or identified. Because there is often a lot of learning associated with the creation of this matrix, it works well to use a spreadsheet and project it on a screen so everybody can see it and update it as discussion continues.

Themes & Classes	Product X	Product Y	Product Z
Administrative			
Municipal boundary lines	☑	☑	☑
Municipal polygons	☑	☑	☑
County boundary lines	☑	☑	☑
County polygons	☑	☑	☑
State boundary lines	☑	☑	☑
Subdivision polygons	☑	☑	☐
Municipal park polygons	☑	☑	☑
State park polygons	☑	☑	☑
National park polygons	☑	☑	☑
PLSS* section lines	☑	☑	☑
PLSS* section polygons	☑	☑	☐
PLSS* township lines	☑	☑	☐
PLSS* township polygons	☑	☑	☑
PLSS* section corner points	☑	☑	☑
Critical infrastructure location points	☑	☑	☑
	☑	☐	☐
Cultural			
Building and structure polygons	☑	☑	☑
Building points	☐	☐	☑
Cultural polygons	☑	☑	☑

Cultural points	☑	☑	☑
Building complex polygons	☑	☑	☑
Transportation			
Curb lines (edge of pavement)	☐	☑	☐
Road polygons	☐	☑	☐
Road centerlines	☑	☑	☑
Road centerlines (dissolved by name and type)	☑	☑	☑
Road centerlines (dissolved by name and block no.)	☐	☑	☐
Railroad Center Lines	☐	☑	☑
Cul de sac points	☐	☐	☑
Hydrography			
Hydrographic area polygons	☑	☑	☑
Hydrographic shorelines lines	☐	☑	☑
Hydrographic lines	☐	☑	☑
Hydrographic center of channel lines	☐	☐	☑
Hydrographic points	☑	☑	☑
Islands	☐	☑	☑
Hydrographic names polygons (names of sub areas)	☑	☑	☑
Hypsography / Terrain			
Contour lines (5 ft interval, 25 ft index)	☐	☑	☐
Contour lines (20 ft interval, 100 ft index)	☐	☐	☑
Hillshade raster	☐	☑	☐
Digital elevation model (DEM) raster	☐	☑	☐
Spot elevation points	☐	☑	☑
Control points (not already in PLSS corners)	☐	☑	☑
Physiography			
Named landform polygons	☐	☑	☑
Topographic Overlays			
Land cover polygons	☐	☐	☑
Vegetation polygons	☐	☑	☐
Map Series Indexes			
Map series index polygons	☑	☑	☑
Ortho Image Base			
Natural color digital ortho image	☑	☐	☐

Figure 2. Matrix showing the data requirements for the example maps. The specific scale or resolution of data is not considered at this stage in creating the matrix. Special case thematic data should also not be considered at this stage. *PLSS stands for Public Land Survey System, which in the United States is used for a large portion of the country as the basis for land management by federal, state, and local agencies.

Data Resolution and Scale

Most municipal or county government organizations that manage base map datasets will have at least two scales or resolutions of products: those products designed for use by people on the ground doing work, and those for use by people managing those people on the ground or directing resources for them. That is, there are products that provide information to people with individual concerns and products for people who manage the information services for those individuals. Depending on the size of the locale, products and services at intermediate scales may also exist.

The initial matrix indicates the basic thematic requirements and geometry types. The next step adds another attribute to the matrix, to indicate resolution or scale (see Figure 3 below). To start, matrix rows were sorted by thematic class then feature class. For each instance where a feature class was used by more than one product, a simple analysis of the scale of use for those products is needed. If more than one scale exists, then an additional row may be required for each additional scale. A new row was tentatively added for each product scale—later analysis may show that some of these rows can be removed if other data proves to be acceptable for use at multiple scales.

Feature Class	**Scale**	**Group**	**Product X**	**Product Y**	**Product Z**
Hydrographic area polygons (1)	1:5,000	A	☑	☑	☐
Hydrographic area polygons (2)	1:25,000	B	☐	☐	☑
Hydrographic shorelines lines (1)	1:5,000	A	☐	☑	☐
Hydrographic shorelines lines (2)	1:25,000	B	☐	☐	☑
Hydrographic lines (1)	1:5,000	A	☐	☑	☐
Hydrographic lines (1)	1:25,000	B	☐	☐	☑
Hydrographic points †	1:5,000	A	☑	☑	☑
Islands (1)	1:5,000	A	☐	☑	☐
Islands (2)	1:25,000	B	☐	☐	☑
Hydrographic names polygons †	1:5,000	A	☑	☑	☑

Figure 3. Expanded product data requirements matrix for the Hydrography theme. The rows of that are marked with a **†** are where the 1:5,000 data could simply be filtered and used on the smaller scale product. The Group column indicates which group or agency within the example county is responsible for the data.

At this point, the highest level of data maintenance for the example county's base map has been captured. The next goal is to reduce the load of maintaining these data. First identify which groups within the organization are responsible for each dataset identified in the matrix of product data requirements. If more than one group appears on the same row, then an opportunity to save effort has been identified. The detailed data requirements for the data in that row should be evaluated to see if redundancy can be eliminated and responsibility can be distributed to other groups. For example, working with the Texas Natural Resource Information System (TNRIS) group in the planning stages of a 1:24,000 topographic map series for their state, they knew they had all the existing DLG road data, and they were assuming it would work well for their topographic product. Early prototypes using these data, however, showed that the data were badly out of date and suffered from numerous feature classification errors (e.g. residential streets coded as interstate highways). A later planning meeting coincidentally included staff from the Texas Department of Transportation (TXDOT) and TNRIS happily ended up using the TXDOT roads data in their maps.

A second way to reduce data needs is to identify automated procedures that can be used to geoprocess data produced by one group such that the transformed data is fit for use by another group. The geoprocessing procedure is a consistently repeatable mechanism that can be used in a timely and managed workflow to create the required data so that the database remains "lean and mean".

Impacts of the Reuse of Data Versus Geoprocessing

The discussion to this point provides the context from which to discuss the overall structure of geographic data as managed within the organization and the implications for product workflows, especially when geoprocessing workflows are involved. A critical consideration is that data used for publication quality maps or near publication quality maps requires product-specific information to be stored in the database to make map production more efficient. For example, in Product Y, the streets are drawn by road type and labeled twice, once by name and a second time with the lowest address number on that block. Figure 2 reflects this requirement; however, since no other maps need the data for "Road centerlines (dissolved by name and block no.)" that data should be produced by geoprocessing it from the "Road centerlines" dataset as part of Product Y's production workflow, rather than stored in the base map.

This type of processing implies that some specific data management or geoprocessing procedures are required for many of the feature classes in a given map. For that reason and also to accommodate organizational work and information flows, a product database is needed. Thus, a high-level view of an organization's base map and product databases that would be used to produce the products described in Figures 2 and 3 might look something like Figure 4.

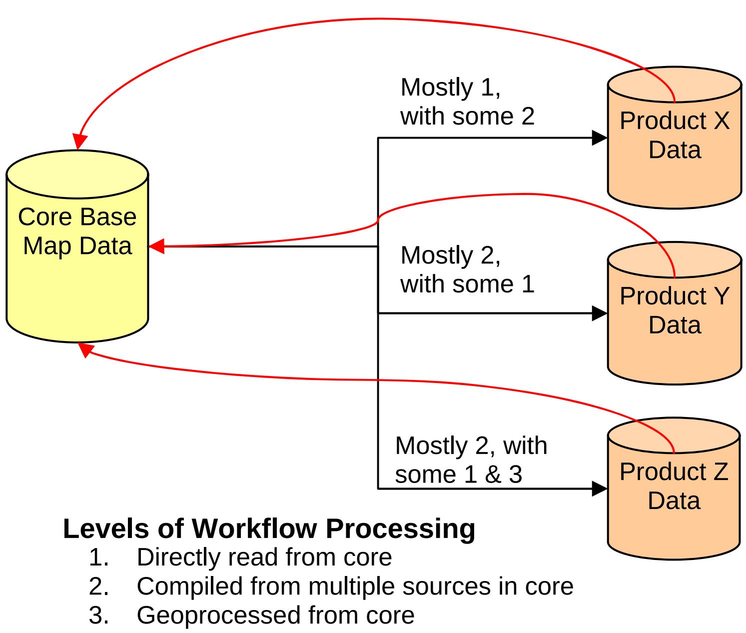

Figure 4. Data flow based on the kind of data processing necessary to produce each product.

Figure 4 shows that several basic levels of workflow exist to get data into a product via a product database. At the first level, directly reading the data from the core base map database is the least expensive. This level should be used if other production considerations, such as access or processing loads on the core database, do not create performance issues for others using the base map data. For instance, if the example county were to decide to create several Web products that used the street centerlines, but displayed them all with the same line symbol, they could directly read the street centerlines from the core database. If performance issues were to arise, a version of that data could be copied to a product database where independent processing would occur.

The second level of workflow incorporates data that is independently managed with data that will be on the map. A good example is place names. It makes sense to manage these data once, to ensure consistency of spelling, capitalization, and abbreviations. Then the names can be joined to the GIS features and exported to the product database in the form of a single feature class. Furthermore, this is a good time to handle product-specific (versus organization-wide) requirements. An example occurs in Product Z, which uses a set of specific abbreviations for park names, e.g., "National Historic Park" is "Nat. Hist. Park".

The third level of workflow involves geoprocessing. This level requires more than simply selecting features that are needed and attaching information to them. If any of a wide variety of geoprocessing operations can be automated, then the burden of storing that geoprocessed data in the core base map is alleviated. The automated process can be performed every time the product database is created. An example of this is the production of contour line data from a Digital Elevation Model (DEM). The DEM may be centrally managed and updated as needed, while contour lines, particularly for larger scale maps, are a resource-intensive dataset to manage—size alone is often an issue). Contour lines can, for the most part, be automatically produced from the most current DEM as needed.

Figure 4 further indicates that there are some important considerations for organizational information flows. The main one from the standpoint of a cartographic workflow is that the mapping process exposes people to the data and therefore affords an opportunity to do heads-up quality checking. Regardless of whether the cartographic staff can directly edit the central database, they should be able, in near real time, to resolve data discrepancies they may find in the process of mapmaking. The implication for a project manager is that if time is only planned for a single automated data extraction processes or geoprocessing operations, especially the computationally-intensive processes, then there is little or no time allowed for errors to be corrected prior to publication deadlines.

Discussion

Considerations for Reusing Data at Multiple Scales

In Figure 3, there are some rows, like Hydrographic Points, in which all three products will use the same data. Reusing data will reduce production costs, particularly if levels one or two of workflow processing are used. Keeping production costs down is obviously a good goal, but cutting costs will at some point impact a product's quality. Identifying and eliminating redundant data collection and production efforts have also become common processes in many larger organizations, such as state or national mapping agencies and many private map producers. Outsourcing data production to places where, in particular, labor costs are lower, has also been a strategy, but the following issue applies in all of these cases. When such processes are conducted in the absence of a product requirements analysis, unexpected reductions in a map's ability to communicate effectively, accurately, and efficiently may result. Such reductions may impact an organization's ability to meet its core business goals. Thus, identifying information themes that are used in multiple products and then comparing the role of that information in each product confirms whether effort is being duplicated.

If separate efforts produce the same or *very similar* information (in terms of the level of geometric detail and types and strategies for storing information in attributes) in several products, then a data reuse strategy should be developed. That is a rather high-brow way of indicating that when the data requirements are the same, choose the better of the two datasets, and when the data requirements are very similar, choose the better of the two datasets and modify its production so it will meet the requirements of all products in which it will be used.

With these thoughts in mind, organizations should have data capture and production processes that produce data to accommodate all of their products in a given scale range. The concept of a scale range is difficult in this context because it is too often assumed that if some of the data for one map is used successfully in another map at a different scale, then all of the data from the first map can be used in any other maps at the scale of the second map so long as other product informational requirements are met.

The idea that data such as road networks, streams, and terrain captured for one map, purpose, and scale will function with equal utility when put into another product at a different scale has proven seductive to many GIS and mapping data managers. By studying each layer of information on many different maps at a relative continuum of scales, it becomes clear that each information theme functions differently and independently with respect to its sensitivity to scale. For instance, a change in the representation of roads from polygons to center lines does not dictate that the representations for hydrography polygons should also change.

Another common assumption is that once a feature has been captured at a sufficiently high level of detail and stored in the core base map, all map products can simply use that feature as is, or as a source for a geoprocessing workflow that derives an appropriate representation of that feature (Timpf and Frank 1995; Davis and Laeder, 1999; Zlatanova, et.al., 2004). Even though much of this work falls into an earlier period of research, many organizations, particularly national mapping organizations, seeking to modernize their map production efforts from still older technology such as tile-based CAD systems still analyze these concepts as potentially viable bases for their modernization effort. The idea that a "golden feature"[1], can be the basis for automated processes to generate cartographically viable representations is untenable for a number of practical reasons. Each of the following four paragraphs discusses a practical reason for avoiding using "golden features" as a basis for geoprocessing and data management workflows.

[1] Scott Morehouse, Director of Software Development at ESRI used that apt term to illustrate that such an approach for deriving alternative representations of features from a single or golden representation as being wishful and unrealistic thinking during an internal meeting several years ago.

Stoter (2005) reports that a number of national mapping agencies as use a "Star" approach, which presumes a centrally managed database from which all cartographic representations of features are derived. Zlatanova et. al. (2004) argue that it should be possible to derive objects of lower resolution from ones of higher resolution. Even when the graphical model for all of such an agency's maps is very similar or heavily reliant upon consistency of convention, the accepted cartographic presentation of the same real-world feature at different scales is often dramatically different at different scales (Buttenfield, 1989). The assumption that different representations are needed in different maps is not a justification for additionally assuming that all representations of the same feature can logically be derived from its most detailed representation (Buttenfield et al 1991, Meng 1997). This is because of the tendencies for:

- geometric and topologic drift (feature shape and size changing to the point where they no longer fit together correctly),
- dimensional collapse (areal features collapsing to lines or points), and
- intransitivities of attribute codes and feature hierarchies.

These tendencies result in creating immense complexity in simultaneously accommodating multiple resolutions of information (Buttenfield, 1995, Weibel and Dutton, 1999, Zlatanova et al 2004) and multiple models of information (Stoter, 2006). For instance, Stoter (2006) contends that the semantic model of the information must also be integrated within the multi-representational management environment. Other intrinsic information models for GIS-based mapping include the GIS model, which is the initial state of a feature's geometric representation and the attributes that describe that feature as well as the graphic model, which is how to symbolize each feature on the map, what order to draw the features on the map, how to place text that further describes each feature, and the tolerances or topological relationships for symbolized features (e.g., how much space must exist between a symbolized building and a symbolized road). The important point is that unless all of these models are taken into account, generalization, or the production of lower resolution representations, cannot be successfully accomplished.

Data capture requirements are often not fully multi-purpose. In working with a number of private commercial mapping companies (who cannot be named without violating non-disclosure agreements), the trend is to buy the best available street and road network data that contains all the informational attributes a given map title or titles require. In working with federal agencies in the U.S., there is wide recognition that no one agency can produce all the data needed to complete a national map. As such, agencies that produce thematically complete maps or datasets, such as the United States Geological Survey (USGS), must depend on other agencies for certain layers of information. The result is cartographic contextual mismatches (sort of like registration errors), and neither fast nor completely successful strategies for trapping and fixing these sort of registration errors has been developed.

Context cannot be predicted; our landscape is not rational. Mapmaking is a response to varied physical and abstract landscapes for the sake of communicating what is there, not a systemization of those landscapes for the sake of mapping. Therefore, the idea of a system for producing and managing multiple geographic representations such as proposed by Timpf and Frank (1995), Davis and Laender (1999), and Spaccapietra et. al., (2000), is suspect because the system will at times impose logic on features that do not fit into the collective description of what is rational. Software developers of the most sophisticated label placement, map finishing, and generalization software routines collectively acknowledge that they cannot make all people happy with their tools and know that their systems must allow for manual intervention to artfully communicate the difficult aspects and intersections among our many landscapes. Another way to approach this to create a system that treats everything as an exception (lacks a main() function); the problem with this approach is that there are simply too many things to map. The British Ordnance Survey, for instance, has more than four billion features in its core database; an exception based system would require too much effort to complete.

Conceptual management of multiple representations of dynamic features is in its infancy. There is a finite scale range through which some representations can, today, be successfully produced with some level of automation. Independent of representational quality, there is a functional relationship between level of automation and the size of this scale range: the higher the level of automation, the narrower the range of scales. The size of that finite range also varies with data theme. For example, natural hydrography has a smaller range than cultural hydrography. Terrain has a very small range, and an administrative boundaries theme has a relatively large range. Furthermore, if the feature is at all dynamic, the rate at which a feature's geometric representation changes will additionally and significantly impact that functional relationship with automation. The effective scale range of use diminishes as rate of change in the feature's representation increases, and the contextual issues in two paragraphs above are also exacerbated.

Commercial GIS software could certainly be improved to better handle production, management, and updating of multiple representations of features that have a common set of properties. Such improvements, even on the cartographic front, will certainly come, if incrementally (Hardy et. al. 2004). From a free-enterprise economics standpoint, though, it is easy to see why commercial GIS software will never have all the dedicated, ready to use tools that just one, much less any organization needs to automate the production of its publication-quality maps from their base map using a "golden feature" approach. Instead, the capability for those organizations to write and integrate tools to accomplish their specific needs is a more commercially tenable direction to take. The informational complexity and high level of graphic finishing most national mapping organizations expect in their maps when considered at the level of detail required to automate production is truly daunting. This is particularly so as it pertains to the inter-feature contextual relationships on a map. Today the best approach, as this paper advocates, is not to get bogged down in finding ways to automatically derive representations, but rather to first start at a high level and understand the basic representational requirements that maps have and use that information to drive an informed data capture methodology to produce or evolve data that are tailored to or fit for use in efficiently producing those maps. For example, occasionally, a group new to mapping brings up the need to generalize contour lines, which ultimately is the most costly and least effective, in terms of data accuracy, way to get a contour line dataset for a smaller scale map. Instead, deriving new contour lines from a DEM or terrain model captured at a coarser resolution would be the better solution. An intermediate solution would be to generalize the DEM; experience has shown that generalizing interpolated data like contour lines magnifies the faults of the interpolation method in the resulting data.

Data Capture Considerations

Contextually informed data capture methods effectively underwrite the successful design of the database and is a prerequisite of base maps. “Informed,” here, means that individual products and programs should not be approached in an isolated manner; rather, an organization’s set of mapping requirements should be analyzed collectively. Each product must use spatially and temporally consistent data. Organizations that produce maps need standards that specifically describe the spatial data accuracy and temporal tolerance of information allowable in their products and data. When organizations do not have such standards to guide their data capture, and they capture data in an isolated fashion, there is a significant risk that their data will not be spatially or temporally integrated. For example, when mapping agencies in the United States, both public and private chose to purchase street centerline data from commercial vendors, they risk their other data, like buildings and structures, not aligning to this independently captured street data.

One important facet of temporally consistent data is that data captured from a single source can be incrementally updated to keep it current. Most local government organizations must be able to conceivably update their data every day. “Snapshots” are often taken of data every so often, such as once every quarter or year, and archived. All data for a given map should either be current or from the same snapshot; mixing data from different temporal snapshots can result in unexpected results and liabilities.

Even with such standards to guide data capture, isolated data capture efforts will erode the spatial integrity in maps and products that must combine data from those separate efforts. Furthermore, when an organization captures the same themes of data at different scales, it is important that the larger scale data be used to inform the smaller scale data, in terms of both the spatial representation and attribution. In fact, an integrated, rather than isolated, data capture regime can significantly improve production efficiency and cartographic quality. A best practice is for related data themes and scales to be captured simultaneously and comprehensively ideally by the same technician. The sequence of data capture is also important; features that are not spatially variable should be captured first, like roads, structures, and boundaries. Features that move or change, such as hydrography and vegetation, should be captured relative to these more stable features. There will always be circumstances that run afoul of rules of thumb, such as boundaries that should follow the center of a river; the main point here is to not capture boundaries and rivers independently and that there is some basic guidance that generally works. In addition, all of this work is based on a control system to which imagery and terrain data are first registered.

One of the common pitfalls in producing a GIS base map is assuming that data for smaller scales can be *(simply)* derived from the data for larger scales through processes of generalization. The reality is, though, that too much information is lost, particularly spatial integration. Data, such as street centerlines, should be captured in conjunction with larger scale data. For instance, an organization might capture pavement and impervious surfaces, which are likely to be digitized at the same scale and similar level of accuracy as that organization’s cadastral data. Thus, the organization builds a polygon data layer for roads and streets. Since it is very difficult to derive a centerline dataset from those polygons that meets typical aesthetic requirements for mapping or analytical requirements for network analysis or address geocoding, the organization should capture at least two street centerline files while digitizing paved surfaces. The first dataset for analytical purposes and larger scale maps would contain centerlines that accommodate medians, dedicated turn lanes and such. A smaller scale centerline file for more general mapping needs would not contain separate lines on each side of medians or visually insignificant connectors and would instead contain the generalized version of the larger scale network.

Management of Scale Independent Data

Another way to reduce the data management workload is to eliminate redundant attribute data processing. That is, when the same process is applied to many or all feature classes in the base map, it is likely that the process can be eliminated through a better data model. Street names, place names, and the specific semantics of how an organization has abstracted the real world into the GIS database are ideally managed once, at the root of the data management regime, rather than multiple times, typically near the end of each product's workflow. For instance, spell-checking and abbreviations should be managed within the core and derived product databases, rather than managed as part of a visual inspection of a nearly finished map. The goal is to ensure that information will appear consistently and correctly on all products and within all analyses thus, eliminating extra steps for the map making or analytical workflows.

Data, like place names, should be maintained separately in the core base map database and joined via a common identifier to other data that are being copied to product databases. But, as noted above, if there are several products that require the same processing of place names, it makes sense to store the results of that process within the core place names dataset. Even though these data are stored in the core base map, they may include consistently used or special abbreviations for common words within place names.

Conclusions

Today, pressure for national scale datasets for the purposes of emergency response and management is exacerbated by content requirements of progressively more refined and specific information. Consistent basic requirements for such data need to be identified and incorporated into local governmental organizations' data capture procedures to ensure consistency and compatibility when their data is ultimately collected by higher-level governmental organizations and published or archived in state or national databases.

While the recommendations in this paper about the form of and processes that characterize good multi-purpose multi-scale GIS base map are an important start, they also lead in some surprising directions. For example, it makes more sense for local governmental agencies to produce and manage data for scales ranging from the highest level of engineering scales that meet local legal mandates for accuracy to smaller scales such as 1:25,000, 1:50,000, or 1:100,000, as needed. The goal is to avoid, when possible, higher level governmental organizations combining and generalizing the local government's products without the local expertise to do so. Problems of accuracy arise when local large-scale data are compiled and used in smaller scale maps *without* local expertise. Higher levels of government either underestimate the cost of properly compiling and generalizing data from local sources, or they do the best they can within the amount of time and money budgeted for the work, under the proviso that the result will be better than doing nothing. The latter is fine in the absence of a plan; thus, the real argument here is for a good plan.

There is no real need for continuous national datasets at large scales. What is needed are well modeled multi-purpose, multi-scale data produced and maintained at the level of local governmental organizations *in federation with* other (especially adjacent) local governments. This is not to say that higher levels of government should not be producing or managing data; there are plenty of areas that have no local agency that can do the work. Higher levels of government should and in many cases already are also serving as repositories and remote and necessarily redundant archives for local data. Higher levels of government also will need to create, by merging the local data, regional datasets that facilitate broader purposes. The need for appropriate and rapid response to large scale disasters of all kinds have dictated such measures be taken. The obvious need to improve the response to such events mandates current, accurate, and well modeled GIS base maps for local scales and purposes that are complete, consistent with their neighbors, and most importantly available. If and once this happens in the U.S., it is vitally important that these data be archived and published for consumption by a host of higher level governmental agencies such as state governments and at the federal level, the National Geospatial-Intelligence Agency (NGA), Federal Emergency Management Agency (FEMA), and the USGS have a variety of missions that require good local data.

References

Arctur, D. and M. Zeiler. 2004. *Designing Geodatabases, Case Studies in GIS Data Modeling.* ESRI Press, Redlands, CA.

Brand, P. 2001. CU engineer: Utility systems under twin towers 'held up remarkably well'. [http://www.news.cornell.edu/Chronicle/01/12.13.01/O'Rourke-WTC.html]. *Cornell Chronicle.* December 17.

Bush, G. W. 2003. Homeland Security Presidential Directive/Hspd-7. [http://www.whitehouse.gov/news/releases/2003/12/20031217-5.html]. Office of the Press Secretary. December 17.

Buttenfield, B.P. 1989. Scale-Dependence and Self-Similarity in Cartographic Lines. *Cartographica,* 26(1): 79-100.

Buttenfield, B.P. 1995 Object-Oriented Map Generalization: Modeling and Cartographic Considerations. In: Muller, J.C., Lagrange, J.P. and Weibel, R (Eds.) *GIS and Generalization: Methodology and Practice*. London: Taylor and Francis, GISDATA Series; Volume 1: 91-105.

Buttenfield, B.P., C.R. Weber, M. Leitner, J.J. Phelan, D.M. Rasmussen, and G.R. Wright. 1991. How Does a Cartographic Object Behave? Computer Inventory of Topographic Maps. *Proceedings, GIS/LIS '91*, Atlanta, Georgia, November 1991, vol. 2: 891-900.

Davis, C. A., and A.H. Laender. 1999. Multiple Representations in GIS: Materialization Through Map Generalization, Geometric and Spatial Analysis Operations. Presented at ACM GIS '99, Kansas City MO.

Hardy, P., M. Briat, C. Eicher, and T. Kressmann. 2004. Database-Driven Cartography from a Digital Landscape Model, with Multiple Representations and Human Overrides. ICA Workshop on 'Generalisation and Multiple Representation', Leicester, United Kingdom.

Meng, L.Q. 1997 Automatic Generalization of Geographic Data. Technical Report of the Institute of Cartography University of Hannover, Germany. 76 pp.

Spaccapietra, S., C. Parent, and C. Vangenot. 2000. GIS Databases: From Multiscale to MultiRepresentation. Symposium on Abstraction, Reformulation and Approximation.

Stoter, J., R. Lemmens, B. Kőbben, and N. Bakker. 2006. Semantic Data Integration in a Multiple Representation Environment. Joint ISPRS Workshop on Multiple Representation and Interoperability of Spatial Data. 24-26 February 2006. Hanover, Germany.

Stoter, J. July 7-8th, 2005. Generalisation: The Gap Between Research and Practice. 8th ICA Workshop on Generalisation and Multiple Representation, A Coruña, Spain.

Timpf, S. and A. Frank. 1995. A Multi-Scale DAG for Cartographic Objects. Auto-Carto 12, Charlotte, NC, ACM/ASPRS.

Weibel, R. and G. Dutton 1999, Generalising Spatial Data and Dealing with Multiple Representations. Chapter 10 in Longley, Goodchild, Maguire and Rhind (Eds) *Geographical Information Systems: Principles, Techniques, Management and Applications.* London: Longman: 125-155.

Zlatanova, S., J.E. Stoter, and W. Quak. 2004. Management of multiple representations in spatial DBMSs. 7th AGILE Conference on Geographic Information Science, Heraklion, Greece, Parallel Session 3.3.