THE CREATION OF A DIGITAL CARTOGRAPHIC DATABASE FOR LOCATOR MAPS[1]

Karen A. Mulcahy
Hunter College

ABSTRACT. ***This paper reports on the procedures and problems associated with the development of a digital cartographic database for the creation of small scale, small size locator maps. These locator maps are used as inserts to accompany thematic and planimetric maps. They are generally very simple outlines of coastlines or political boundaries, but may also contain boxed or shaded areas, graticules, and occasionally, thematic content. Several cartographic datasets were evaluated and three were chosen for use to cover a scale range of 1:800 million to 1:30 million. The datasets were then reviewed and edited. Finally, guidelines were prepared for the use of these datasets in an ARC/INFO environment, with particular emphasis on generalization and smoothing.***

This paper reports on the procedures for developing a digital cartographic database for the production of locator maps. The project was conducted for the National Geographic Society but the principles developed are generic to other situations. Locator maps are an especially important tool of geographic reference. National Geographic magazine articles make extensive use of such maps. Locator maps help the map reader find the area depicted in the page map they accompany. As the National Geographic Society integrates automated techniques into the mapping process a need has developed for digital cartographic databases to support locator map production.

The small scale, small size, locator maps such as those used in National Geographic magazine may be just simple outline of continents such as South America or Africa. Due to the mandate of the National Geographic to "increase and diffuse geographic knowledge" it is essential that the area depicted on a page map be clearly located in its spatial context. The maps may contain boxed or shaded areas, graticules and, occasionally, thematic content as well. The purpose for developing a database is to make locator map production more efficient.

The linework and shading for most page maps are initially prepared using ESRI's ARC/INFO software program. Each page map is currently developed as a customized dataset digitized for the particular map. The creation of an accompanying locator map is approached in several ways. The linework needed for the locator map can be extracted from one of several databases saved from previous work which may have an appropriate level of generalization. Linework may also be extracted from the current customized dataset being used and transformed to the needed level of generalization. Another possibility is to digitize a locator map at the appropriate level of generalization. There are problems with each of these approaches. Extracting linework from previous datasets can be a hit or miss operation based on the knowledge of individual operators, on both the datasets available and on the generalization capabilities of the software. This approach often requires several cycles of editing which consists of work on the map by an operator and corrections suggested by map editors. Digitizing a new locator map is also time consuming and would still need to be examined by the map editor and corrections made. The result of these methods is that locator map creation can be as time consuming as the production of the larger, more complex page map the locator map is to accompany.

---

[1] **Winner of the Student Paper Competition, Middle States. This research was supported by the National Geographic Society Internship Program.**

The raw material available for the development of a database were several previously digitized datasets of global extent, most of which were digitized at National Geographic. These datasets ranged in the level of generalization from a dataset consisting of only coastlines to one consisting of more complex coastlines, political boundaries, rivers and lakes.

Database development can be approached from one of two directions. In the first approach a collection of individual locator maps could be produced. Some of the most often used locator maps are outlines of Africa, South America, Australia and India. Global locator maps in an orthographic projection are also used frequently. There are several problems with this approach. Global locator maps use the same data and projection in many cases but each map is centered on a different point on the globe. This would require producing and saving a multitude of global locator maps centered at various points. Locator maps consisting of regions rather than a specific continent or country, like the global locator maps, are difficult to predict. Even creating outlines of the larger countries and the continents for general use may be impractical because the level of detail for any one locator map may change. A map with drop shadows would require less detail than one without them. An examination of past examples shows that very similar locator maps have differing amounts of detail.

The second approach to locator map database development consists of selecting one or more cartographic datasets of global extent. Each locator map is extracted from the global dataset as needed. Global locator maps can be centered at any point on the globe, regions can be of any size or extent, and individual locators can be created with any amount of detail required. The locator database project took this approach.

For a locator map database consisting of data of global extent, the existing datasets and available generalization techniques needed to be investigated to determine whether a single, detailed dataset could be used for all locator maps or if several datasets would be needed to cover the wide scale range of locator maps. During this project several datasets were evaluated and chosen for use based on accuracy of linework and possible scale range. These datasets then needed to be edited and recommendations for their use developed.

Evaluation of the Datasets

The first stage of the database project was to evaluate the cartographic datasets. Several datasets of global extent at varying levels of generalization are available. Of these, three datasets are stored on two different computer platforms using different software programs. The WORLD projection package runs on a PC computer and the ARC/INFO package runs in a networked workstation environment. Software programs are able to store cartographic data in a geographic grid system of latitude and longitude coordinates and both programs can transform the data to different kinds of map projections. The WORLD software package is specifically designed for the purpose of transforming data from latitude and longitude coordinates to various map projections. It provides quick results with relatively little operator training. The ARC/INFO software program is a full scale geographic information system which has map projection capabilities as a small section of the overall program. Although it is very powerful, ARC/INFO requires extensive operator training and is very time intensive when compared to the WORLD software package for performing map projections.

The WORLD projection software was used for the initial evaluation of the cartographic datasets. To produce the map projections, small batch progams were written using a system editor and executed to transform data to the desired projection and to display it to the computer screen. Alternatively, the data can be sent to a pen plotter or saved as a file which can be exported to other software packages. Figure 1, Western

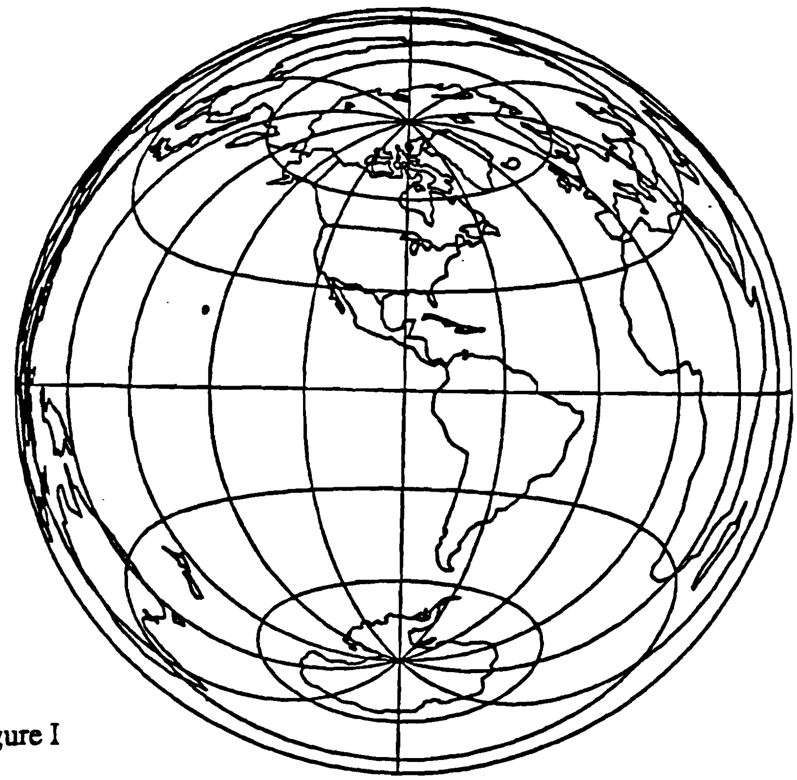

Figure I

Western View

View, was produced using WORLD and transferred using a conversion program then printed using a laser printer.

Because locator maps can range in geographic extent from a global view all the way to small states such as Rhode Island, the scale range for this project was identified as between 1:300 million and 1:30 million. An extensive series of plots of the cartographic datasets at various scales was produced and the proposed scale ranges for two datasets were developed. A third dataset, which was only stored on the workstations using the ARC/INFO software, was tentatively chosen to fill in the largest part of the scale range.

The simplest dataset in terms of line complexity and features was one which was originally included with the WORLD software package and consists of very simple shorelines. This dataset was named WORLD300 because it was identified to have a useful upper limit of one to 300 million. A locator map of global extent in an orthogaphic projection at a scale of one to 500 million is about the size of a quarter. Of greater complexity is a dataset identified to be useful at scales between 1:300 million and 1:75 million and is named WORLD75 to represent the one to 75 million upper scale limit. This map was originally digitized from a National Geographic world map at a nominal scale of 1:100 million and consists of shorelines and political boundaries. A third dataset, WORLD30, chosen to cover the scale range between 1:75 million and 1:30 million, consists of shorelines, political boundaries, and lakes and rivers, was also digitized from a National Geographic world map but at a nominal scale of 1:30,840,000.

The choice of these three sets of digital cartographic data constitute the initial stage of the database project. The quality of the data was known because these datasets had been used previously. The next stage in the database project was to present the datasets in hard copy form to the map editors for corrections and to further investigate generalization techniques which may be required even within the proposed scale ranges.

Editing and Generalization

The three datasets were each projected and plotted onto paper for editing. In order for each area on the globe to be examined by the map editors, each dataset was projected to five views: Western, Africa, Asia, Arctic and Antarctica. (See Figure 1 for the Western View employed and Figure 2 for the Arctic view.) The Lambert Azimuthal Equal Area projection was used so that between the five views every area on the globe is represented as nearly conformal. The Lambert Azimuthal equidistant was not used because the software would not project all five views. The datasets were then plotted onto paper using an electrostatic plotter with as thin a line width as possible to try and mimic actual line weights used in the magazine.

The purpose of plotting onto paper was for the map editor to make corrections to the linework. Figure 3 shows the plotting scheme for the three datasets. Each dataset was plotted in each of the five different views at its largest and smallest proposed scale. Although not shown here, WORLD300 was also plotted at a smaller scale. The largest scale plots of each dataset is the scale at which it was edited. The smallest scale plots of each dataset were plotted for comparison purposes between the datasets. Depending on the amount of detail desired one dataset may be chosen over another where the scale ranges meet. These plots of the three datasets were turned over to the map editors for changes. The editors were able to edit at the largest scale while looking at the smallest of the scale range for reference. The map editors marked the plots for changes which corrected or removed political boundaries and which changed, added, or deleted shorelines and islands in order to make the entire dataset conform to a consistent level of generalization.

The level of generalization for each dataset was directed toward the smallest end of the scale range. The two larger scale datasets, WORLD75 and 30, will need in many cases further generalization when used at a scale at the lowest end of its scale range. Three types of generalization will be considered: line simplification, area generalization and displacement of features. This is a practical breakdown of generalization types for this project. There is no computer software that can do all types of generalization automatically. Each type needs to be considered separately.

Displacement of features is required when a reduction of scale brings features so close together as to hide the relationship between them. Two islands may be each considered important to represent on the map due to a strait between them. It may be necessary to exaggerate the space between the land masses in in order to depict the water passage. Currently the only realistic means of performing such a displacement of features is to produce the locator map at the desired scale and send it to the map editors for changes.

The area generalization of features in the locator database project refers to the inclusion, exclusion, or even the creation of a new feature which represents the amalgam of islands as scale changes. As the scale of a map is reduced some islands are considered so important in their spatial context that they need to be represented even if they are actually smaller land masses than other islands which will not appear on the map. Alternatively, several islands that appear to merge at a reduced scale would be best represented by one island that is an amalgam of several islands. The only practical available method for treating the amalgamation or elimination of features during automated generalization is limited to a coding scheme which would allow certain objects to drop out and possibly for other representations of features to be included as scale is reduced.

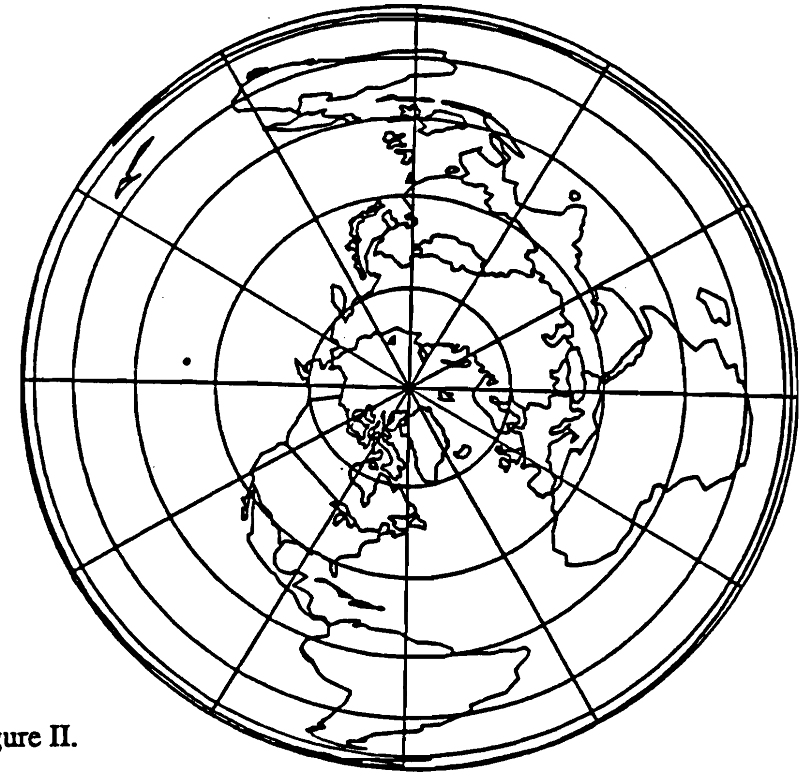

Figure II.

Arctic view

Similar to the displacement of features, area generalization in terms of the reduction or merging of features will need to be done interactively with the map editors. There are area generalization methods being researched but they have not been developed as much as line generalization methods have. Brassel and Weibel (1988), in their review article of automated map generalization, mention methods of automated area generalization such as line generalization algorithms, expansion and contraction operations, and a shape smoothing method using the medial axis transform. The problem of area generalization is more difficult to solve than that of line generalization. Much of the research and success in automated map generalization has been in the area of line generalization and not area generalization. This is especially evident when using computer software. Line generalization algorithms are very common but area generalization techniques are absent. More research and development of these techniques is required before they can be implemented and accepted on a wide scale.

Two techniques for line generalization were used: a spline function and the Douglas-Peucker algorithm. A spline function reduces or increases the number of points by use of a tolerance value. Reducing the distance between points with a spline will add points to the cartographic data, smoothing it, while increasing the distance between points with the spline will reduce the number of points and thereby generalize, or simplify the line. Common uses of the spline are to clean newly digitized data of redundant data points, or a spline can be used to smooth lines after simplification for an improved aesthetic quality (McMaster 1989).

The Douglas-Peucker algorithm is used to reduced the complexity of a line for a map at a reduced scale. It has been thoroughly tested and is widely accepted as an effective

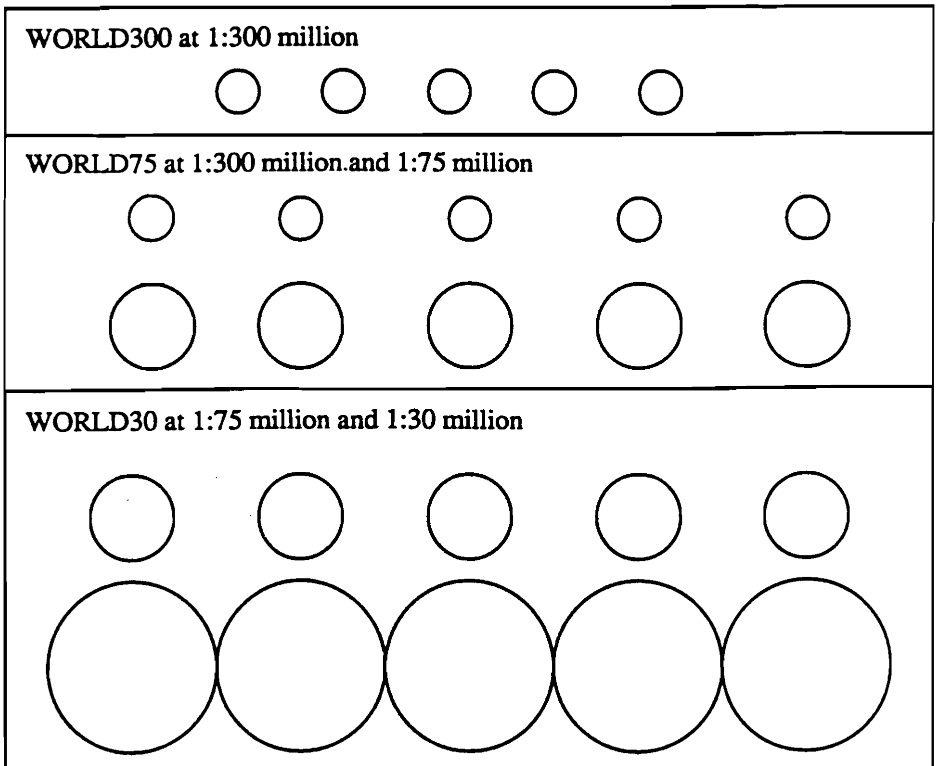

Figure III. Plotting scheme for editing the three datasets. (Not to scale.)

---

method for line simplification. Testing of this algorithm on the datasets was conducted to establish the relationship between scale reduction and a tolerance value for defining the level of simplification. This may have been possible if only one dataset was being used but in this case three datasets were being examined. Tolerance values related to actual line width were converted into map coverage units and tested. The results from each dataset were quite different even when reducing scale by similar factors. Further testing needs to done using this algorithm. The final step of the locator database project was to prepare recommendations of which dataset to employ for a particular locator map. A chart was prepared which suggests an appropriate dataset for global, continental, country and state locator maps. The chart includes all of the continents, countries and the United States, and as such, was too large to be included here. A few points can be made about the chart. Two suggestions were made for several countries which fall on the scale boundary between two datasets. The choice of one over another would depend on the amount of detail required for a specific locator map. The chart also lists a large number of small countries in the other column, indicating the limits of the scale database, WORLD30. The locator database should eventually be expanded to provide cartographic data for these small countries and for US states and for provinces in large countries such as Canada.

Conclusion

The problem of multiple representations from a cartographic database is an ongoing area of research in digital cartography. For the production of locator maps, datasets were evaluated and chosen for use within a structure of a single large database containing multiple datasets of global extent. Scale ranges for the datasets were chosen and the editing needed to prepare the datasets for use begun. A brief investigation of automated map generalization methods was undertaken and recommendations were made as to which of the datasets should be used for various locator maps.

In terms of providing a more efficient means for locator map production, the locator database project provided a base upon which to continue work. Testing of generalization methods is needed and new methods for increasing the efficiency of locator map production should be explored. The locator database should also be expanded to provide a source for all locator map needs. For example, another dataset, or sets, should be included for locator maps of small countries, states and provinces.

The locator database project can also be examined from the point of view of a first step in planned database development. A comprehensive database would increase productivity in the computer-generated map production process. At present nearly every page map is produced from a custom database developed for that job. The locator map database can be viewed as a prototype for further database development. Many issues dealt with in the locator project are generic to any digital database design. Generalization techniques are the most difficult to evaluate and implement. Issues of accessibility and storage requirements, not addressed explicitly in this project, would also be very important in further overall database development.

References

Environmental Systems Research Institute (1987). ARC/INFO Software Users Guide (Redlands).

Brassel, K., and R. Weibel (1988). "A Review and Conceptual Framework of Automated Map Generalization" International Journal of Geographical Information Systems 2: 229-244.

Douglas, D., and T. Peucker (1973). "Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature" Canadian Geographer 10: 110-122.

McMaster, R. (1989). "The Integration of Simplification and Smoothing Algorithms in Line Generalization" Cartographica 26: 101-120.

Shapiro, R., and C. Martin (1989). "Challenges and Solutions in Map Publishing" Technical Paper. Fifty-fifth Annual Meeting, American Congress on Surveying and Mapping pp. 237-245.

Voxland, P. (1987). WORLD Projection and Mapping Program: Version 4.20 (Minneapolis: University of Minnesota).