QUALITY CONTROL AND STANDARDS FOR A NATIONAL DIGITAL CARTOGRAPHIC DATA BASE

Francis J. Beck
Randle W. Olsen
U.S. Geological Survey
510 National Center
Reston, Virginia 22092 U.S.A.

ABSTRACT

The identification, development, and implementation of quality control procedures and standards for digital cartographic data within the U.S. Geological Survey have been accelerated, and shortcomings in existing quality control procedures have been identified through increased use of the data. Data quality concerns are directed toward six areas; header record integrity, labeling of source and lineage, positional accuracy, file completeness, topological integrity, and attribute accuracy. The USGS, National Mapping Division, is developing a series of Technical Instructions describing standards, procedures, and data use.

U.S. Geological Survey is responsible for chairing the Federal Interagency Coordinating Committee on Digital Cartography. The Committee has representatives from 23 Federal agencies. The USGS also supports and is represented on the National Committee for Digital Cartographic Data Standards, which operates under the auspices of the American Congress on Surveying and Mapping.

INTRODUCTION

The U.S. Geological Survey (USGS) conducts a formal program for developing, implementing, and updating standards related to digital cartographic data. In addition, the USGS has taken a lead role in coordinating standards activities throughout the federal government and is a strong supporter and contributor to local government, private sector, and national standards activities. The National Mapping Division (NMD) of the USGS is continually establishing or refining standards which directly affect the design, content, and quality of the National Digital Cartographic Data Base (NDCDB). The ongoing standards and quality control program for digital cartographic data is being shaped by data user requirements and feedback, improved digitizing technology, in-house data requirements, and non-USGS standards-setting activities.

Publication authorized by the Director, U.S. Geological Survey.

CURRENT DATA STANDARDS

The U.S. Geological Survey has been investigating the gathering and processing of digital cartographic data since the early 1970's. The process of defining and populating a data base of digital cartographic data has been active for approximately 10 years. The current data base contains a collection of digital data organized by scale-specific, product-specific, cartographic units.

The design phase of the data base required many basic decisions pertaining to data source materials, accuracy requirements, attribute coding methodologies, levels of structuring, digitizing techniques, processing systems, formats, and distribution procedures.

Source Materials

The standard graphic products of the USGS were selected as the fundamental source of digital cartographic data. This selection was based on the widespread availability of these graphic products (approximately 90 percent coverage of the continental United States), the large investment in this mapping program, the inherent accuracy of the graphics and their suitability for diverse applications. The 7.5-minute 1:24,000-scale quadrangle series was selected as the primary source for digital data. The 1:2,000,000- and 1:100,000-scale series were selected later as sources for building small-scale data bases.

Level of Structuring

Defining a level of topological structuring for the data was a difficult and critical decision made early in the program, resulting in a decision to incorporate a high level of topological structuring. That is, explicit references to node-area-line relationships are gathered and kept in the digital fields. This decision was based on the idea of making the data compatible with advanced data manipulation and analysis systems such as those having current widespread popularity in the geographic information system (GIS) environment.

Accuracy Requirements

A data accuracy policy was formulated once the decision was made to use standard USGS graphic products as digitizing source materials. It was also recognized that the digital data would be later used to prepare revised graphics. The digital data were to accurately represent the source graphic and therefore have its implied accuracy and level of content. Therefore, the stated policy has been that the data in the NDCDB has a level of content equivalent to that shown on the source graphic (for any given data category) and is positionally accurate to within ±0.005 inches (0.13mm) of the location indicated on the source graphic that itself was compiled to meet U.S. National Map Accuracy Standards.

Attribute Coding

The NMD attribute coding scheme was developed in an attempt to reach two primary goals. The first was to gather the symbolic information displayed on the source graphic through an orderly, understandable, and easily applied set of numeric codes. The second goal was to anticipate, as much as possible, the types of data analysis questions that would eventually be asked of a highly structured file of digital cartographic data. The coding scheme was further defined to accommodate digitizing individual categories, or subcategories, of information from the graphics and to satisfy the requirement to gather nodes, areas, and lines as input to the structuring software. A seven-digit coding scheme was designed consisting of a three-digit major code (with two digits specifying the data category) and four-digit minor code which identifies the data element type (node, area, line) and provides specific details of the feature digitized. The leading digit of the minor code is normally zero. If non-zero, it is used to provide specialized information about a feature. Table 1 describes the DLG major and minor attribute coding scheme and assumes a leading zero in the minor code.

DATA QUALITY ISSUES

The NMD currently is directing a significant level of effort toward six areas related to digital cartographic data quality — header record integrity, labeling of source and lineage, positional accuracy, file completeness, topological integrity, and attribute accuracy. Generally deficiencies in the data that have been identified by users of the data have been related to one of these six areas.

Header Record Integrity

The header records of data in the NDCDB contain many elements of quality related information, such as, file name, date and scale of source materials, ground reference system and map projection parameters, data resolution, and coordinates describing the geographic domain of the file. Most of these items were defined during the design of the NDCDB; however, a free format concept for many of these values resulted in serious limitations on the ability to automatically search for specific data elements or to perform a rigorous quality control on header record contents. Errors in the header record were primarily caused by keying mistakes during data entry. Current format specifications designate specific fields for all these data elements. Prompting routines of present structuring software specify items to be entered and a broad screening

Table 1.—U.S. Geological Survey Digital Line Graph Coding Scheme

Major Code	Base Category
020	Hypsography
050	Hydrography
070	Surface cover
080	Non-vegetative surface features
090	Boundaries
150	Survey control and markers
	Transportation systems
170	Roads, trails
180	Railroads
190	Pipelines, transmission lines, miscellaneous transportation
200	Other significant manmade structures.
	Non-Base Category
300	U. S. Public Land Survey System
Minor Code	For All Categories
001-099	Node
100-199	Area
200-299	Line
300-399	Single-point (degenerage line).
400-499	General-purpose codes (apply to multiple feature types).
600-699	Descriptive codes

of the information is performed to verify its appropriateness for a specific field. Additional software checks are made prior to data base entry to verify or build as many elements as possible, such as, sheet name, scale, and coordinates.

Labeling of Source and Lineage

Source and lineage information for data in the NDCDB is found in the header record of each file and in documents available for each type of data, such as, digital line graph (DLG), digital elevation model (DEM), land use and land cover (LULC), and geographic names information. Several source and lineage items were mentioned earlier in "Header Record Integrity." Documentation describing the graphic and digital NMD programs provides information on map compilation procedures, accuracy and content requirements for standard graphics, current digital data gathering techniques, and information on the software used to process the data.

Positional Accuracy

Horizontal and vertical accuracy of data in the NDCDB is dependent on the data source. In the case of DLG's the horizontal accuracy is within ±0.005 inches (0.13mm) of the position indicated on the source graphic. This level of digitizing accuracy is intended to retain the accuracy of the graphic. Positional accuracy is currently verified by visually comparing high-accuracy photo plots of the data with the stable-base source material.

Current data gathering procedures for DEM's result in different vertical accuracies. The accuracy of data gathered on the Gestalt Photo Mapper using automatic image correlation or by manual profiling from stereomodels is highly dependent on the resolution, contrast, and scale (normally 1:80,000) of the source photographs. Two accuracy levels of DEM data are maintained in the data base, one level containing data with a RMSE of up to 7 meters and the other having a 7-15 meter RMSE. Two additional sources of DEM's are contours digitized during map compilation (and subsequently gridded into a DEM) and digitizing contour separates of existing maps (either by semi-automatic line following or automatic scanning). The accuracy of these data sets expressed as a root mean square error is required to be within one-half contour interval of the source graphic. Initial editing and testing of DEM's are performed within the production unit. Test points are obtained from a 7.5-minute quadrangle map of the area or are developed during the aerotriangulation phase of the mapping. A final quality-control step for DEM's is performed on a DEM editing system. The system has the ability to display the DEM in color or black and white, and the display may be a shaded relief model, an anaglyphic stereomodel, color-coded elevation zones, or a histogram of all elevations. A graphic corresponding to the area of coverage may be registered to the DEM through an attached high-accuracy digitizing table, and this permits editing and accuracy testing of the DEM. All DEM's are viewed, edited, and accuracy tested on this system prior to entry into the NDCDB or, for data already resident in the data base, prior to distribution.

File Completeness

The basic tenet of the NMD concerning file completeness is that the digital file for a given category of data will contain the same level of detail shown on the source graphic. Map information is generally portrayed by the position, symbol, and (or) label of a feature. Often all three are needed to accurately convey this information.

Therefore, completeness has implications on the positional representation of a map element, such as x,y coordinates, as well as the attributes attached to that element. Both positional and attribute requirements have been addressed by current NDCDB data gathering and processing techniques. Text data gathering requires further attention. A limited number of feature labels are gathered in the form of numeric codes as part of DLG attribute coding, such as, route designations and widths, various located symbol labels, culture feature descriptors, land grant names, and origins of the Public Land Survey System. However, text for named features is not accommodated under current DLG digitizing procedures. A related activity within the NMD is the development of the Geographic Names Information System (GNIS) which uses the 7.5-minute graphics as the basic source for its text information. Current research in the NMD is directed toward an enhanced DLG data structure that will directly accommodate textual information. Investigations are also being made into the possibility of merging text data from the GNIS with the appropriate elements of DLG data.

Topological Integrity

Topological integrity, as with file completeness, has implications on both positional and attribute content of the data. Gathering and retaining node-area-line relationships must be positionally and qualitatively accurate. Lines must begin and end at nodes and no lines may intersect (cross) without the presence of a node. Left and right areas must be indicated and appropriately coded. Much has been written on the characteristics of topological structuring and this paper will not attempt to redefine all requirements of a structured file. Current topological integrity quality control is performed primarily by automatic checks made during data processing through the structuring software. In addition, manual error checks are made against listings, plots (both black-and-white as well as color raster), and a variety of color displays. These latter checks are made on interactive editing systems and are used to detect and correct topological inconsistencies in the data.

Attribute Accuracy

Attribute codes, or feature codes, are relied upon to convey much of the information gathered during the digitizing process. Unfortunately, coding of data is often the most subjective phase of the digitizing process and requires human interpretation of the source graphic and the coding scheme. Consequently, a first step in achieving uniform and accurate attribute coding is to provide clear, comprehensive instructions on how to apply each attribute code. Following several years of operational experience, the NMD has adopted and published DLG attribute code standards which give specific instructions on applying its coding scheme. Further operational experience shows that additions, modifications, or clarifications to these attribute codes are needed; however, this document forms the basis for all DLG coding.

Automated and manual methods are used to check attribute codes prior to entry into the NDCDB. Software routines verify that a major code is appropriate for the data category in which it is used. Each code number (major and minor code) is automatically checked against a master list to see if it falls within the acceptable range of numbers for a specific element type of a data category, for example, 050 0101 (reservoir) is verified as an acceptable area attribute within the hydrography data category. Some codes are defined as parameter codes in that the numbers of the minor code explicitly convey a special significance, such as, indicating a water level, a Federal Information Processing Standard (FIPS) State or County code, or the elevation of a survey marker. These codes are verified against a list of acceptable values such as FIPS codes or a range of numbers, or checked as to type of entry, such as numeric or alphabetic. Manual checks are made by comparing specific numbered elements and their codes against the corresponding element on the source graphic or previously prepared coding overlay. Data are also interactively displayed by code, in color or black-and-white, to verify coding and file completeness. Color and black-and-white raster plots are also prepared for specified code combinations to check coding accuracy.

STANDARDS ACTIVITIES

The Branch of Technical Management (RSTM) within the NMD has the overall responsibility for developing, implementing, and maintaining standards or procedures documents relating to the National Mapping Program. All such documents comprise the NMD Manual of Technical Instructions. Specific documents included in this series are standards, procedure manuals, software documentation and users manuals, and data users guides. There are three organizational sections within RSTM responsible for documents related to digital data, graphic products, and automatic data processing. The standards setting process extends to other USGS Divisions (Geologic, Water Resources, Information Systems) through the Data Standards Committee. This Committee acts on suggested earth science data standards from the Divisions within the USGS or from other agencies. A document accepted as a USGS data standard will normally be submitted to other standards groups, such as the National Bureau of Standards for FIPS Standards, and the American National Standards Institute (ANSI), for their consideration as a standard.

The NMD is continuing to refine digital cartographic data standards and quality control procedures to support its digital cartography program. The application of these standards and procedures extends beyond the Survey's map-making activities. There is significant impact on the map user community as well. Consequently, the USGS has taken a lead role in defining and implementing digital standards that are expected to have impact throughout the government and private sectors.

A major effort within the federal government to coordinate digital cartographic data standards is underway in the Federal Interagency Coordinating Committee on Digital Cartography (FICCDC). This Committee was chartered by the Office of Management and Budget to reduce the duplication of effort by federal agencies in digitizing map data and developing geographic data files and is chaired for the Department of the Interior by a representative from the USGS. The FICCDC has established a Standards Working Group specifically to facilitate data interchange and thus avoid unnecessary costs for data conversions.

This working group has examined existing exchange formats and is developing a specification for a Federal Geographic Exchange Format (FGEF). The FICCDC Standards Working Group is working very closely with the National Committee for Digital Cartographic Data Standards (NCDCDS) which is under the auspices of the American Congress on Surveying and Mapping and includes membership from Federal, State, and local governments, as well as private companies. The NCDCDS has a working group on data organization which is looking at the FGEF and its applicability to a national data exchange standard. It is possible that the FGEF will be one of several formats to be accommodated by a national data exchange format being defined within the NCDCDS.

CONCLUSIONS

Since the beginning of the digital cartography program the USGS has made major advances in gathering, processing, verifying, and archiving digital cartographic data. Sales of these data have steadily increased and these additional users have provided feedback which shows that careful attention must be paid to both the application of standards and quality control procedures. Therefore, the USGS is giving added emphasis to these aspects of their digital cartography program. The results of this activity will be particularly evident in the management and technical support provided to federal, national, and other standards groups.

SELECTED REFERENCES

Allder, William R., and Elassal, Atef A., 1984, Digital Line Graphs from 1:24,000-Scale Maps: U.S. Geological Survey Circular 895-C, 79 p.

Corbett, J. 1979, Topological Principles in Cartography. U.S. Bureau of the Census, Technical Paper 48, 50 p.

Domaratz, Michael A., Hallam, Cheryl A., Schmidt, Warren E., and Calkins, Hugh W., 1983, Digital Line Graphs from 1:2,000,000-scale Maps, U.S. Geological Survey Circular 895-D, 40 p.

Fegeas, R.G., Claire, R.W., Guptill, S.C., Anderson, K.E., and Hallam, C.A., 1983, Land Use and Land Cover Digital Data: U.S. Geological Survey Circular 895-E, 21 p.

Peucker, T. and Chrisman, N., 1975, Cartographic Data Structures. The American Cartographer 2, No. 1, p. 55-69.

Peuquet, D., 1984, A Conceptual Framework and Comparison of Spatial Data Models. Cartographica 21, No. 4, p. 66-113.

U.S. Geological Survey, 1985, Standards for Digital Line Graphs, Part 3, Attribute Coding: U.S. Geological Survey National Mapping Division, Manual of Technical Instructions.

White, M., 1983, Tribulations of automated cartography and how mathematics helps. Automated Cartography: International Perspectives on Achievements and Challenges (Auto-Carto VI), Volume i, ed. B. Wellar. Ottawa: The Steering Committee for the Sixth International Symposium on Automated Cartography, p. 408-418.