LESSONS LEARNED IN ESTABLISHING AUTOMATED CARTOGRAPHIC DIGITAL DATA BASES

William T. Riordan
Deputy Director, Management and Technology
Defense Mapping Agency
Washington, D. C. 20305

It is generally agreed that establishing an automated data bank is both complex and expensive. It is expensive because the digitizing, editing, and evaluation of the digital data is still very labor intensive. It is complex because of the large number of items that must be considered from full scale digitizing systems to specific data items to be collected. It is this high cost which encourages us to make maximum use of the digital data that has been collected. We feel that greater utilization of the data can be achieved through multiple product data bases. While we have taken some steps in these areas, we still have a long way to go. I will be discussing four basic areas:

a. Definitions for both data and data bases

b. Standardization

c. Data formats

d. Quality control

In a sense, the data base is not different than the office file. For the file to be useful, we must know what we have (definitions), how good it is (evaluations), and where it is (organization)so we can retrieve data when we need it.

Definitions

In any new endeavor we must start out with a definition of terms so that we can communicate. This need is especially true in the data base area. We must define not only the data elements, that is the individual items which are stored in the data base, but also the different types of formats and coding structures. As you learn you become more discerning and the definitions become more specific.

Data Elements

The main data elements are the features; that is, roads, streams, etc., to be digitized. One really needs a data element dictionary in which each element is described in unique and unambiguous terms. Both users and producers must have a clear understanding of each element. We must be able to distinguish between ponds and lakes or streams and rivers.

For example: In one of our management systems we had the data element called significant date which was related to map currency. One DMA Production Center used the date that source material was collected, the other Center used the date that the map was ready for printing. The definition was made more specific to distinguish between the two dates. In our terrain elevation format, there is a data element for a sequential block count. The definition did not specify starting with a 1 or 0, and as you might expect, one Center used 0 and the other used 1. The definition has since been modified to indicate starting with zero.

Format

The format is the positional order of data elements in a record and records on a tape. Here again it is very important to eliminate any possible ambiguity in the format definition. To exchange digital data one must be able to read another producers' tape. Most tape reading problems are related to a misunderstanding of the format. We recommend the generous use of examples in the format definitions and the performance of actual tests with potential users before publication of the final format.

Coding Structure

To conserve storage space, we never spell out the names of features, but rather use some type of abbreviated code to represent each feature. There are many coding structures in use today; hierarchial, sequential, alphabetic, and random to name a few. The particular structure is not critical. However, the structure must allow the coding to be comprehensive and expandable for future additions of features. In our preliminary cartographic feature standard, we used a hierarchical structure. The single character representing types of features, such as culture, hydrography and hypsography only allowed for ten types of features. This does not provide for an adequate number of feature types and will be expanded in the next version of the standard. We will add features such as soils, landforms, land use for terrain analysis.

Standardization

Dedicated Resources

Standardization is at best a slow process and if not diligently pursued, it never happens at all. Standardization is frequently viewed as an unnecessary constraint that requires potentially disruptive change. Typical responses to statements about standardization are: it will cost more money; or it won't work. Therefore, a very important first step in standardization is to select some personnel who will be dedicated to the definition and implementation of standards. It is a difficult task and will take considerable time and effort to accomplish. The personnel selected must have a positive attitude towards standardization and must also have a broad knowledge of the production organization to maintain the proper perspective. We have established positions in our Headquarters and in our Production Centers specifically for this purpose.

Resolution of Conflicts

Because of the changes required by standardization, one must recognize that differences of opinion and conflicts will arise. Therefore, it is also necessary

to establish review and approval channels that allow resolution of these problems. Top management participation in the review and guidance process is essential to provide authenticity and to insure acceptance of the standards. Where changes are required, they must be well planned to minimize the impact on people and production. The training and education of managers and production personnel is essential to maximize the acceptance of change.

Degree

Standardization must be accomplished judiciously. One must constantly be reminded that the goal is to provide more cost effective operations, not to over-standardize which may do just the opposite. One of the biggest problems is to decide what to standardize and to what degree. In the ever-changing digital environment we must standardize those items which will tend to remain fixed with time and to a degree which aids but does not constrain the utilization of data for the community of users. This has been a particularly perplexing problem to DMA since we have multiple production organizations that have different equipments and processes which tend toward nonstandardization in themselves.

Evolution

Our approach is to introduce standards as an evolutionary process, through new development and procurement of equipment rather than the potentially disruptive and costly process of retrofitting. As new systems are developed, standard data elements and formats can be used at virtually no extra cost or effort. This was done with the introduction of the terrain elevation standard into our production line. We developed software to produce a standard cell of data and also conversion software to generate products from the standard format. It is now possible to generate the terrain elevation standard format and then the product. This was accomplished without any disruption to the production process.

Data Formats

The use of the terrain elevation standard for both exchange and storage has worked out reasonably well for DMA. However, discussions with other users and producers in the digital community indicate that some are not willing to adopt the standard for internal use. These discussions have lead to a better understanding of the different producer environments and the realization that we must be more discerning in describing the type and use of standards and specifications. We really need to define three types of specifications: exchange, storage and product. Each of these specifications has a special role to play in the overall production and user environment and, therefore, certain characteristics which are important. Each specification must consider format, content and accuracy. The exchange specification should be oriented to serve the total community for ad hoc requests. The storage specification should be oriented to serve internal needs for the multiple functions of update, production and exchange. The product specification is designed to serve a group of users with a predefined set of data.

Exchange Specification

If the exchange specification is to serve the total community of users for ad hoc requests, its format must be flexible enough to allow for a wide range of data types and data combinations. It should also allow for the cell size, resolution and reference systems to be variable so that it can satisfy many different applications.

The specification must also define a catalog of standard elements to uniquely identify all the elements and to make their availability known to the community. The exchange specification would include the definition of a format and a catalog of data elements and codes but would not define any specific data content. Data to be exchanged would depend on the requester. This format would not only satisfy ad hoc requests by system developers, but also periodic requests by coproducers. If each producer of digital data develops conversion software from their storage formats and codes to the

exchange formats and codes, then they can all exchange data and make maximum use of all available data. The DMA cartographic feature standard will be modified to include all types of features and to serve as an exchange format.

Storage Specification

Storage specifications for internal data bases should be oriented toward production. The formats should be designed to maximum reformatting for the three uses of update, production and exchange. Although the exchange format can be used for storage, normally the number of update and production operations will exceed the number of exchange operations and a special storage format will be desirable. Recognizing that product requirements can change, the storage formats should be flexible enough to allow for additional features as the needs arise. However, the specific design of format and codes is left up to the producer and is not constrained by the exchange format. The storage specification also must define the standard data sets to be stored in the data bank.

The use of different storage and exchange specifications provides a highly desirable isolation between each user and the community. Users may change and upgrade their internal systems and storage formats with no impact on the community. As long as they maintain conversion software to and from the exchange format all internal operations are transparent to the remaining users.

Product Specification

Product specifications are oriented toward a group of users who have agreed upon a specific format, content and accuracy. The digital landmass system (DLMS) is a typical example of a product specification. The amount of data to be generated must be of sufficient quantity to justify the development of the specification. If only a small amount of data is to be generated, it should be satisfied through the general exchange format rather than as a product. If a new product specification is to be developed, it is very important to determine the data requirement to a high

degree of specificity so that costly changes in published specifications are avoided.

Quality Control

The normal approach to provide a quality product is to provide in-process and post-process quality controls. For maps and charts this is relatively straightforward since the various overlays can be viewed directly as they proceed through the production line. Digital data is different in that it is not directly viewable. The data must be plotted on a CRT of paper to be viewed. This means that the functions of editing, verification of content, and evaluation become more complicated than for the graphic process. Batch processing on the computer followed by a line plotter to edit and verify data is simply not adequate. The amount of accounting effort required to keep track of the various tapes and the potential for error is just too high. In the collection phase, interactive CRT systems are essential for cost-effective operations. We are gradually converting our digitizing stations to be interactive. It is a good rule that a unit of work, such as a predefined cell or area, should remain in the interactive environment until fully qualified. It can then return to the batch environment for routine reformatting and transformations.

Frequently the interactive operations are performed on dedicated computers for a local production department and the batch operations are accomplished on a central main frame computer. For planning purposes, it is important to distinguish between the two types of operations and to determine the computer requirements for each product. The computer requirement for each product can then be used along with the production program and the volume of each product to forecast the total computer requirement. However, product requirements can change so the computer requirements must allow for flexibility and expandability.

In addition to providing quality control during the production of the digital data, it is equally important to maintain adequate process control in the copying and shipping operations. We have had more than our share of problems with shipping tapes with the wrong

data or incorrect labels or seven tracks instead of nine tracks. So much time and effort is spent in qualifying the data that we tend to overlook that last step of getting that data to the user properly identified and in a form he can use.

Summary

If we have learned anything, it is that the establishment of automated cartographic data bases is a never ending task. With the everchanging product requirements and changing technology we will always be in a state of transition. Therefore, we must allow for change and plan for a smooth transition which will have minimum impact on internal production as well as the external community of users. The organization of the data bank must allow for the coexistence and communication between old and new formats, systems and technology. We feel that the development of an exchange specification for all types of geographic feature data is an important next step for us and for the digital data community, and we will be working with other digital data producers during the next twelve months to establish this specification.