ABSTRACT

METHODS FOR COMPARING DATA ENTRY SYSTEMS

by
Don Hansen, FWS
Mike Dwyer, BLM, CSO
Marilyn Mogg, BLM, DSC

A need to standardize many automated geo-processing capabilities has prompted the U.S. Fish and Wildlife Service (FWS) and the Bureau of Land Management (BLM) to evaluate the systems currently in use within the two agencies. The Automated Digitizing Systems (ADS) and the Analytical Mapping System (AMS) are geographic data entry systems developed by BLM and FWS respectively for later processing by the Map Overlay and Statistical System (MOSS). ADS and AMS are similar in many respects in that they produce comparable digital map products. The methods used to generate those product are, however, generally very different. These methods were analyzed in terms of their inherent efficiences and deficiencies. The performance of each system was evaluated on the basis of several benchmark tests which addressed user time requirements, computer loading factors and data accuracy. The results of all analyses, in addition to responses to a user survey, were incorporated into recommendations for improvement of each respective system. The implications associated with using ADS and AMS in a mapping program were evaluated in terms of established cartographic standards.

A task was initiated in 1984 through a joint agreement by the U.S. Fish and Wildlife Service (FWS) and Bureau of Land Management (BLM) to study and compare the capabilities and characteristics of geographic data entry systems used within the Department of the Interior (DOI). The primary objective of the study was to identify those components of data entry systems that contribute most significantly to their ability to compile high quality geographic data bases. Information collected during this comparison would be very useful in efforts to standardize geoprocessing techniques throughout DOI. The purpose of this paper is to report the methods that were used to evaluate the Automated Digitizing System (ADS) and the Analytical Mapping System (AMS).

The collection and analysis of large volumes of spatial data is a major effort of most land and resource management concerns. Traditionally, the inventory and planning tool has been pertinent data plotted on a physical map. This, however, can be cumbersome to work with, especially when complexity magnifies with additional themes and layers of data. Geographic Information Systems (GIS) provide a facility for the manipulation of large volumes of digital spatial data without the limitations encountered with a physical map. Both FWS and BLM are involved in extensive database construction and cartographic modeling projects to assist managers with their land use decisions. Digital databases can be accessed easily to obtain inventory data and manipulated in some logical method to generate useful information. Most spatial data is not, however, immediately available in digital form (an exception to this is digitally recorded, remotely sensed data). Spatial data must typically be manually transcribed onto a physical map incorporating geographic control. Lines are to describe the locations of features such as roads, or the boundaries between areal features, such as vegetation communities. The problem then becomes the transformation of a physical map to a digital map compatible with a GIS. This transformation is achieved by manually digitizing map lines with the assistance of specialized computer hardware and software packages.

A map digitizing system consists of several components each contributing an important function to the whole process. A certain hardware configuration is necessary to run the program. Programming techniques and user documentation affect the maintainability and usability of a system. Data storage file structures impact computer resources, as well as maintainability of established databases, and interfaces with a variety of GIS's permits transportability of databases. The data entry process itself, can be further categorized into 6 distinct stages of sub-processing encountered by the operator in the following order of occurence:

Project Set-up
Map Registration
Data Capture
Editing
Polygon Formation/Verification
Database Management

These ten components/processes were analyzed with the goal of addressing four topics of primary importance as identified by users of geographic data entry systems:

- Accuracy of the geographic database
- Efficiency of the data collection process
- Impacts on the computer environment
- Interface between the program and it's users

Accuracy of the Geographic Database: The results of a cartographic inventory or modeling effort are only as accurate as the data that were used. Land use decisions involving legal boundaries necessitate strict controls over data throughout its processing. Three types of analyses were performed to identify sources and causes of error introduced to the database and to compare the significance of program generated error with that introduced during the original map drafting process.

First, the sub-processes that potentially impact accuracy of data were analyzed to determine what programming procedures are being used. For instance, one of the major sources of error can be attributed to map registration which establishes the link between physical coordinates of the digitizing tablet and geographic coordinates of a map projection. The link is a set of mathematical expressions derived using sampling methods. Polynomial regressions or geometric relationships are used to best approximate or "rubber-sheet" a spherical coordinate system using input from a flat map. The algorithms used to do this are characterized by inherent deficiencies that can be analyzed mathematically.

The data capture stage of digitizing introduces the primary source of human error to a database. Digital map quality is dependent on how accurately the operator traces map lines with the cursor. However, other factors such as digitizing mode (point, increment or stream coordinate measurement), weeding of points to reduce data volume, and "adjustments" to coordinate locations made automatically by the program, can also contribute significantly to the accuracy of the final product.

Polygam formation/verification and editing are other sources of error, as again, adjustments are made to the data by the program and/or operator. Giving the operator a large amount of editing flexibility could also allow further degradation of the data originally collected during data capture.

After reviewing the methods used in the computer programs to retain accuracy, a test was conducted to measure the amount of error being produced during the registration process. This test consisted of measuring the deviation of a digitized point from the true geographic location taken from a map. Maps representing 3 map projections, 6 representative scales and 4 geographic regions were selected. The maps were generated on stable medium with a grid of tic marks located at known latitude/longitude coordinates using a GIS map plotting function.

A set of digitizing tablet coordinates (in inches) were entered into the registration programs of each respective data entry system, thereby eliminating human biases from the test. After the maps were registered, coordinates (in inches) of the internal tics were likewise entered into each system using options that return the transformed, geopgraphic location (in latitude/longitude) of the points. The returned latitude/longitude of each sample point was compared to the latitude/longitude that was expected and the deviation (in seconds of latitude and longitude) was computed. The sampling design permitted statistical analyses of the effects of various combinations of projection, scale and geographic region on accuracy of the database. An example is the strong correlation that was found between map scale and accuracy.

Based on the results of this test an analysis of the "cartographic implications" of error being introduced to geographic databases was performed. The intent was to assess the significance of documented errors from the test in terms of impacts on cartographic products developed from these data bases. Arc second deviations were converted to map inches at various scales. The difference in performance of the two systems was analyzed in terms of significance levels, and a regression was computed to describe the relationship between scale and accuracy.

Efficiency of the Data Collection Process:

A second set of performance tests were used to measure the amount of operator effort required to complete the digitizing process. Factors that influence efficiency include volume and frequency of entering information at the data entry terminal, the logical sequence or digitizing lines and entering polygon attributes, and digitizing mode.

Processes that are implemented automatically by the program require less operator involvement than operations carried out manually. Besides the initial capture of lines, editing capabilities and procedures can contribute significantly to the efficiency of the data collection process.

A test was initiated to compare data entry systems in this respect. The procedure consisted of digitizing maps of varying levels of complexity and measuring the time necessary to complete each step of digitizing. Three levels of map complexity were tested. Low complexity was defined as containing a moderate number of straight lines requiring a minimum number of X, Y coordinates. A public land survey (PLSS) map was chosen to represent low complexity. High complexity was defined as containing a high density of extremely irregular lines bounding many polygons. A 1:24,000 scale map of vegetation cover was chosen to represent high complexity. A third map delineating the locations of large range allotment boundaries was chosen as moderate complexity for its few number of moderately irregular lines.

Experienced data entry operators were selected to digitize the maps. Strict controls were implemented to ensure uniformity of operating conditions throughout the testing. Time was accrued each time the operator logged onto the computer and stopped accruing when logged off. Each time the operator moved to a different stage of processing, time spent on the processor was recorded. Functions such as polygon formation, which did not require operator involvement, were not measured for this test. (Statistics were collected for use in analyzing impacts on computing environment. See below.)

Results of this test permitted analysis of each digitizing process separately. Insight was gained into the potential for enhancements to specific procedures that required input of unnecessary or redundant data.

Impacts on the Computing Environment:

In addition to the clock time that was being recorded during the digitizing test, measurements were also being recorded on variables that indicate the demand on the computer's processing resources. Some data entry tasks are inherently "CPU intensive," demanding dedication of the computer's processor to complete the function. Polygon formation is one such task. Volume of calculations as well as the amount of data that must be transferred between terminal, memory and disc storage severely affect computer response being felt by other users. Programming techniques can be used to maximize the efficiency with which data is handled. Variables measured to evaluate the programs in this respect included: "CPU seconds" (central processor time dedicated to a specific task) and number of I/O calls (requests the processor to input or out put data to devices). As in the time test, CPU statisitics were measured for each processing stage. The resulting analysis provided the basis for critiquing programming techniques used in the two data entry systems.

Interface between Program and Its Users:

This topic addressed the so-called "user friendliness" of a computerized process. It is generally preferred to obtain computer programs that require a minimum amount of training to become proficient. This is especially important with large and complex geographic data entry systems. Factors that contribute toward usability include: user documentation outlining speicific procedures to be followed; "online documentation" available for quick reference; clarity of online menus and instructions; and functional capability. Powerful editing functions, for example, not only minimize time requirements but may also lower the frustration associated with highly intricate work. Program documentation is essential to a system analyist attempting to maintain large volumes of program source code necessary to operate one of these systems. All these factors affect the user's overall ease in using the system. Therefore, the analysis included frequent consultation of users of the two systems. Two surveys were written and distributed to operators and programmer personnel. Responses to the survey provided insight not normally available or utilized by developers of systems such as these. The conclusion quickly becomes evident that the end user should play an important role in the development and enhancement of geographic data entry systems.

Conclusions:

This study not only provided FWS and BLM personnel objective evaluations of two data entry systems, but also relied on a systematic framework for the evaluation of other similar systems. Additionally, the knowledge obtained by comparing two established systems generated many ideas that can be used in future development and enhancement efforts.

Donald B. Hansen
Remote Sensing Specialist/Systems Analyst
Technicolor Government Services, FWS Operations
P.O. Box 9076
Fort Collins, Colorado 80525 Tel. 303/226-6183

Marilyn A. Mogg
Cartographer
Bureau of Land Management
Division of Surveying and Mapping Systems
Denver Federal Center, Bldg 50 D-417B
Lakewood, Colorado 80215 Tel. 303/236-0112

Michael F. Dwyer
Cartographer
Bureau of Land Management
Colorado State Office
1037 20th Street
Denver, Colorado 80202 Tel. 303/844-3066