COMPUTER-ASSISTED ENVIRONMENTAL DATA HANDLING: ISSUES AND IMPLICATIONS

Kenneth E. Gordon
Department of Geography and Regional Planning
Western Washington University
Bellingham, Washington 98225

Introduction

This paper is a report of a study of computer-assisted spatial data handling systems. The study's focus is upon systems which store, manipulate, or display topographic, land resources, or similar environmental data in digital form. The study seeks answers to questions which would be of greatest interest to users of environmental data, especially those considering the use of computer-assisted techniques or the use of computer-stored data from other sources. Issues for spatial data handling and system design are identified based upon an observation of the collective data handling and data use requirements of different user groups and the potential applications and limitations of different types of spatial data handling systems. Data handling encompasses the following operations: 1) data acquisition; 2) changing data to useful formats; 3) storing data in or on some medium; and 4) retrieving and manipulating data for display and analysis. Spatial data handling systems include data-base management systems, cartographic systems, geographic information systems, and various special hardware and software configurations (e.g., for Landsat data processing). The common characteristic is the capability of the system to store and process spatial data in a manner that both the data attributes and their geographic location can be retrieved.

Many public agencies, research institutions, and private corporations are experimenting with and utilizing computer-assisted spatial data handling technology. The inventory of what applications may be performed, what problems may arise, and what is desired by the user are basic preliminary steps for information system design. This paper briefly describes a research study designed to answer these and other similarly important spatial data system design questions, and reports some of the findings of the investigation (1).

Research Technique

A survey of system users was conducted during the summer of 1978. Its population was limited to spatial data handling systems operating in or supplying data for the Pacific Northwest states of Washington, Oregon and Idaho (2). The investigation was designed to be both descriptive and analytical. Many aspects of data handling were queried in the survey: system application; data handling capability (software); method of data acquisition; data storage, manipulation, and display. Many characteristics of the data were also of concern. These included scale, coordinate reference, precision and resolution, source, and form of location identification. Other important issues included: data and software transferability; factors inhibiting the wider application of systems; differences between actual and desired data types, software, and applications; and the extent and characteristics of digital data coverage. A mailed questionnaire, followed by in-person or phone interviews, was the method chosen to conduct the survey (3).

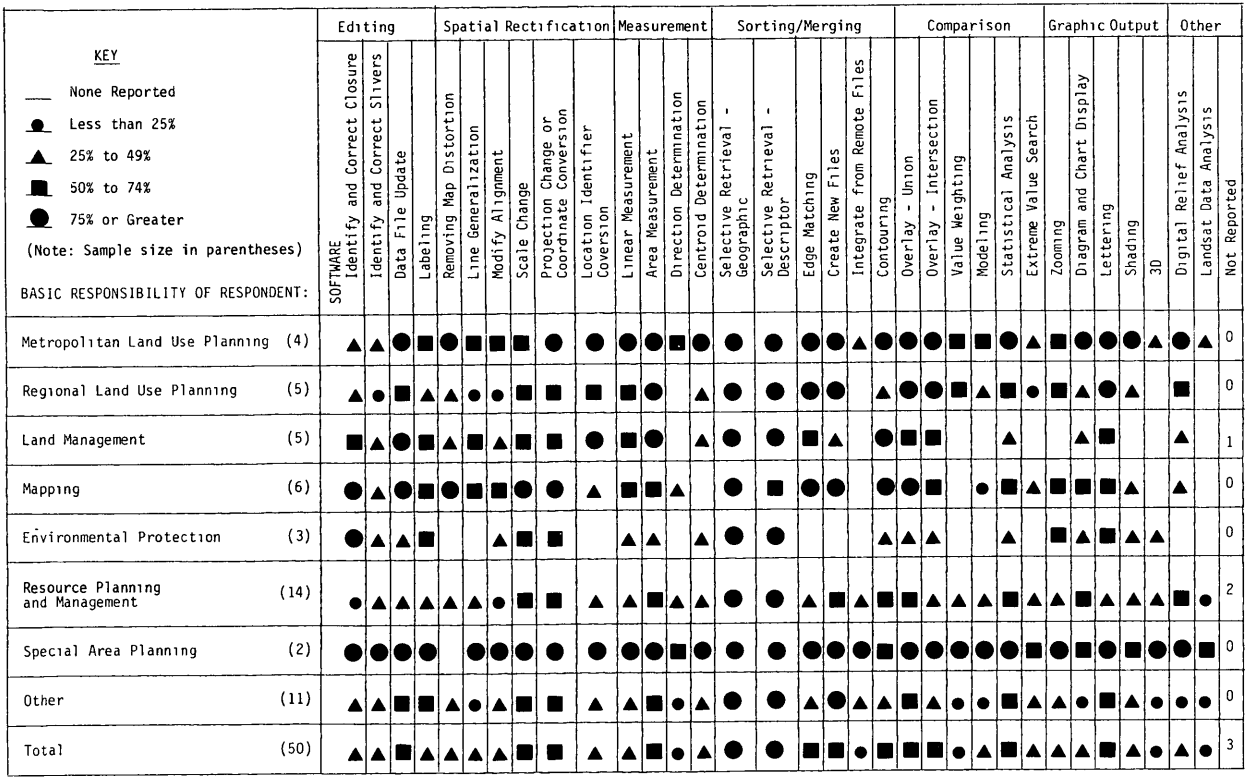
The survey population consisted primarily of planning, natural resource, and environmental agencies, and the spatial data processing systems of the agencies. Data was acquired from thirty-nine federal systems, ten state systems, five regional systems, four municipal systems, and six corporate systems. The chosen sample of fifty responses was quite diverse. Some of the agencies and systems included in the survey were the following: Environmental Protection Agency - STORET; Soil Conservation Service - Advanced Mapping System and Natural Resources Data System; Geological Survey - Digital Mapping Systems, Computerized Resources Information System (CRIB), Geographic Information Retrieval and Analysis System (GIRAS), and WHATSTORE; Washington State Department of Natural Resources - Gridded Inventory Data

System (GRIDS) and Calma Mapping System; Oregon Department of Revenue - Computer-Assisted Mapping System (CAMS); and many municipal government, and corporate systems - Puget Sound Council of Governments, Lane County, Oregon, City of Tacoma, Battelle Northwest Laboratories, Boeing Computer Services, and Weyerhaeuser Corporation.

More than thirty individual characteristics of the systems, the data, and the agencies' use of the systems were recorded from the questionnaire. The responses were coded and keypunched and then tabulated, using the CROSSTABS option of the Statistical Package for Social Sciences. The program counts the number of pair-wise comparisons between selected variables, and prints out tables of the frequency and percentage of correlation. Thus, for example, the number of times respondents reported both aerial photography as a data source and a particular size of coverage could be tabulated. Cross-tabulation was performed on twelve data characteristics such as source, precision, resolution, scale, and location identifier. Significant correlations were found between many of these variables, and many expected correlations were not observed. The sample also could be sorted by any other chosen descriptor. A profile, albeit from a limited sample, was thus developed for the characteristics of the total sample, federal versus non-federal systems, eleven different system types, and eight different system applications. It was thus possible to differentiate the data use, geographic referencing, or data handling capabilities and perceived needs of regional planning agencies from municipal planning agencies, of environmental protection from natural resource management agencies, of data-base management systems from true geographic information systems, or grid referenced systems from point, line, or irregular polygon referenced systems.

A sample tabulation is included. It portrays the desired data handling capabilities of groups of system users. Comparison with a similar tabulation of the operating characteristics of the systems facilitates the assessment of unmet user need. This type of assessment provides, by empirical description, an overview of the limitations and potentials of different types of systems, and of the preferences and unmet needs of different types of data users.

COMPARISON OF THE DATA HANDLING SOFTWARE DESIRED BY EACH TYPE OF SYSTEM USER



Results

Space does not permit the reporting of the findings of the research at the level of comparative detail of the system and use profiles. However, some significant insights were developed based upon analysis of the responses from the total population. Also reported are some observations based upon a comparison of the tabulated responses for federal and non-federal systems.

General observations

A number of empirical, analytical, and deductive observations were made in the course of this study. They reflect some of the problems and potentials of the use of computer-assisted spatial data handling systems, and provide comment upon the status of geoprocessing in this country.

Stage of Development - The systems are in various stages of development. Most are operational, but still improving their data handling capability. The systems built around particular data storage and retrieval functions are least likely to be considering new applications.

In-House Programming - A very high percentage of the respondents reported that the data handling capabilities were developed in-house. This procedure is especially true of the non-federal systems, and has caused problems of lack of documentation and lack of concern for data or software transferability. Vendor-supplied software is available for nearly every data handling task, but few agencies take advantage of this source of software. The only software which was nearly universally supplied by the vendors was the graphics component.

Limited Application - Most systems are dedicated to the performance of very specific applications and are constructed around the performance of these tasks. Though many systems seem to have the software to perform more sophisticated data analysis and display, there are few reports of systems being used to their potential.

Data Accuracy and Data Documentation - There is a noticeable lack of concern for data accuracy and data documentation. Less than 15 percent of the respondents were aware of the precision of their data, and most do not maintain descriptions of the basic characteristics of the data necessary to assess its utility. The lack of data documentation is alarming. Access to and analy-

sis of the data are thus limited by the lack of general knowledge of where the data came from, how recent is its vintage, who collected it and how, to what degree it is generalized, and how it is interpreted.

Data Integration - Systems are able to store many different types and formats of data, each uniquely referenced by geographical location. Most respondents report that they can and do commonly store environmental data with non-environmental data such as land use, census, facilities, political boundaries, etc. The extent of logical integration of the different data files is not known. Half of the respondents report the ability to change scale, change projection, and convert between different encoding formats.

Diversity of Design Options - The characteristics of the systems and the way in which data are handled in the systems are very different even among respondents with similar administrative responsibilities and data handling needs. There are therefore many different system design options which may satisfy similar user needs.

Most Desired Software - The greatest proportion and frequency of unmet spatial data handling needs are reported for the following: Landsat data use and analysis, value weighting, direction determination, shading, overlay, projection change, centroid determination, edge matching, and statistical analysis. The most common types of data handling software include these: selective retrieval of geographic and attribute descriptors, creation of new files, lettering, overlay, scale change, coordinate conversion, area measurement, and contouring.

Factors Limiting Expanded System Use - Limited mandate, budget, and time are reported to be the predominant factors restricting the greater application of systems. Data availability, accuracy, and reliability are not perceived to be very limiting. There is greater desire indicated for more hardware and software to process data than to improve the data itself. The lack of trained personnel is also a significant deterrent.

Interrelationships of Data Characteristics - The data descriptors which were found to be related to one another, and therefore factors which deserve added attention for system design, are the following: data type and data source, resolution and size of coverage, location identifier and size of coverage, scale and resolution, scale and size of coverage, and scale and precision. The size of the area and the scale of the

encoded data appear to be very significant. The factors for which correlations are implied by deduction, but for which analytical examinations failed to show correlation, are these: data type and location identifier, resolution and data source, data source and size of coverage, precision and size of coverage, and location identifier and resolution.

Comparison of Federal and Non-Federal Systems

The federal systems normally are designed around large data bases of primary data. Some include applications programs to assist the data users, while others are simply designed for data storage and retrieval. Non-federal systems are usually designed around broader data use objectives and can handle ad hoc inquiry. Federal systems are more likely to be developed by vendors, while non-federal systems are likely to be in-house programmed. The size of the area of coverage, and the primary nature of the data in the federal systems has a significant bearing upon the characteristics of the data within the system's data base. This may influence the ability of the federal systems to supply useful data to other digital data users and accounts for the perceived need for data conversion software. An assessment of the characteristics of the data shows, for example:

1. Original data sources are very different. Federal systems contain data primarily from field monitoring stations and field surveys. Non-federal systems contain data primarily from published surveys and maps, conventional aerial photography, and from pre-encoded data sources;
2. The predominant encoding scale of data from federal systems is 1:62,500 to 1:1,000,000. The predominant scale of data in non-federal systems is 1:24,000 or larger;
3. The predominant location identifier of data in federal systems is coordinate point. The predominant location identifier of data in non-federal systems is irregular polygon, with significant percentages of coordinate point and grid identifiers as well;
4. The predominant coordinate reference for federal systems is latitude and longitude, while the predominant coordinate reference for non-federal systems is State Plane Coordinates;
5. The predominant map projections for federal systems are Transverse Mercator and Lambert Conformal Conic, while the predominant map projection for non-federal systems is Polyconic.

Conclusion

There is a time in the evolution of any new technology when it is appropriate to step back and look critically at the experience to date. In this case, a research technique was designed to benefit from the recording and comparison of the unique characteristics of systems designed to date for handling environmental data. The intent was to provide aggregated assessments of various characteristics for future reference. The sample was small, but the results are believed to be representative. The research resulted in the recording of many illuminating observations about geoprocessing systems and their use. A researcher can only make these observations, pointing out areas which seem to deserve attention. The responsibility for using the results of these and similar research efforts in a constructive manner rests with the people making data handling decisions.

Notes

1. The study was performed as a master's thesis at Western Washington University.

2. The survey coincided with other studies of the Land Resources Inventory Demonstration Project, a program of data and technology application involving federal, state, and local agencies in the Northwest. Partial funding was granted under University Consortium Interchange No. NCA2-OR862-801 from the NASA-Ames Research Center.

3. Documentation of the survey content and conduct are contained in a report prepared for NASA entitled, "An Investigation of Digital Geographic Data Handling Activities and Digital Environmental Data Coverage in the Pacific Northwest States," 1978.