User Requirements for Framework Geospatial Data

Steven M. Frank, Michael F. Goodchild, Harlan J. Onsrud and Jeffrey K. Pinto

Abstract: *Common sets of geospatial data, usable across many applications, have been proposed as a method to promote GIS data sharing. Questions arise as to the appropriate characteristics of framework data. Which features need to be included in framework data sets? What accuracies are required for features? Which geocoding schemes are needed? How often do framework data sets need to be updated to remain useful? While much anecdotal information is available, studies and surveys on geospatial data needs have been primarily limited to hardware/software issues or confined to a particular state or region. The National Center for Geographic Information and Analysis (NCGIA) conducted a nation-wide mail questionnaire survey to gather information concerning the technical requirements for geospatial data. The questionnaire targeted existing users of GIS or GIS products (i.e., maps, reports, etc. generated from GIS). These users were asked for responses regarding their data needs for that class, including content, tasks for which the data are used, format, geocoding scheme, positional accuracy, vertical accuracy (if needed), updating interval, needs for historical data, and the sources for data currently being used. The NCGIA mailed out 1360 questionnaires; 595 usable questionnaires were completed and returned. The returned information was analyzed across sectors of government, private industry, and academia by geographic region and by professional area of application, showing the technical preferences for framework data sets across each sector profile.*

There is a general agreement among the users of geographic and land information systems that common sets of geographic data that users could adjust to their own particular data needs would promote greater data sharing among the various players in the GIS community (NRC 1994). These would present a significant cost savings to federal, state, and local government agencies and an economic stimulus to American industry by reducing the costs of collecting and processing useful digital geographic data. Such common data sets—often referred to as "core" or "framework" data sets—could be collected by designated government agencies or by participating members of the private sector and added to the public domain, via the National Spatial Data Infrastructure (NSDI), for the use of any and all interested parties.

Steven Frank is an assistant professor in the Department of Surveying at New Mexico State University. He received his Ph.D. in surveying engineering from the University of Maine. He is also a registered land surveyor in New Mexico and California.

Michael Goodchild is professor of geography at the University of California, Santa Barbara, and director of the National Center for Geographic Information and Analysis.

Harlan Onsrud is chair of the Department of Spatial Information Science and Engineering at the University of Maine and a research scientist with the National Center for Geographic Information and Analysis.

Jeffrey Pinto is an assistant professor of management in the School of Business at Penn State–Erie.

Within the U.S. geographic information community there has been protracted debate in the past two years over the value of various data sets, and their importance in the National Spatial Data Infrastructure. A range of criteria for prioritization have been suggested, and strong arguments have been presented for the importance of the various types of data, in addition to the framework function described earlier. This paper is not concerned with resolving these arguments, or even with establishing the relative value of various types of data to the user community. Rather, our objective is to help to clarify the debate by identifying the specific aspects of data sets that are perceived to be of most value to specific communities of users.

To help discover which framework data sets should be given priority for incorporation into the NSDI, the Federal Geographic Data Committee (FGDC) has funded a mail questionnaire survey that was conducted by the National Center for Geographic Information and Analysis (NCGIA) with oversight by a peer focus group. The results of this survey are intended to provide a guideline for the selection and/or prioritization of possible framework data sets into the NSDI. In the space available we can only offer an overview of results by summarizing the responses. Those interested in greater detail are encouraged to obtain the full NCGIA technical report (NCGIA Technical Report 95-1). The report as well as the survey response data files are available through the NCGIA ftp site (ncgia.ucsb.edu) under the same report number.

Several existing studies have been done in this area. Ventura (1991) and Sandberg (1992) present survey findings of local government GIS uses in Wisconsin and Michigan respectively. Budic (1993) has done the same for local government GIS uses in the Southeast. Other literature in this area seem to be regional studies and concentrate on local government uses. These studies also address many organizational and implementation issues which were not of interest to our study. However, such literature proved extremely useful for developing a base upon which to design our study.

Goal

The goal of this survey was to select the technical criteria that may be used to identify and prioritize the framework data sets for the National Spatial Data Infrastructure. These framework data sets may be identified by defining the technical specifications required by the users of these data sets, including content, tasks for which the data are used, format, geocoding scheme, positional accuracy, vertical accuracy (if needed), updating interval, needs for historical data, and the sources for data currently being used.

It should be recognized that geographic information users may be required to use certain data not because of their technical merit, but because of political or organizational expedience. Organizations may have agreements to share or purchase geographic data from predetermined sources, thus forcing the use of certain types and qualities of data. This questionnaire did not investigate possible political or organizational factors.

It is expected that no single set of criteria will be appropriate for all GIS users, but that identifiable sets of criteria will emerge that can form the basis of selecting current digital geographic data sets that best meet such criteria as NSDI framework data sets and for improving such data sets to meet the full needs specified by the different subsets of criteria.

Survey Data Collection

A peer focus group reviewed the questionnaire, the survey sampling strategy, and the survey results. It was the intent to keep the focus group small and to keep discussions on materials up to date. Focus groups members were encouraged to consult with others outside the focus group and to forward comments or suggestions from others where appropriate. Members of the focus group were: Donald F. Cooke, Geographic Data Technology, Inc.; William J. Craig, University of Minnesota; Charles Dingman, U.S. Bureau of Census; Cliff Kottman, Intergraph Corp.; David Mark, State University of New York, Buffalo; Mike McDermott, U.S. Geological Survey; Gerard Rushton, University of Iowa; Nancy von Meyer, Fairview Industries; Nancy Tosta, FGDC; and Michael Domaratz, FGDC.

Data were collected by a questionnaire sent and returned by mail. The questionnaire allowed data collection from a large group of people in a relatively short time at a relatively inexpensive cost. The questionnaire was designed to capture background information useful to correlate and aggregate data into meaningful units of comparison and to capture information on the technical details of digital geographic data used or needed by the respondents.

The unit of analysis for the questionnaire was an individual using geographic information. The term "geographic information" as used here is intended to include information about the spatial locations and the nonspatial attributes of objects on or near the Earth's surface. Collection of questionnaire data at the individual level allows aggregation of data by many different means, such as discipline, organization, experience, etc.

Other possible units of analysis considered were organizations, networks, and episodes of geographic information use. Aggregate use of geographic information by an organization or network is difficult to quantify and compare and would not allow the depth of data collection available at the personal level. Episodes of geographic information use may be frequent in many cases, making them difficult to track other than by automated methods. Such methods are not available for the breadth of uses intended to be covered by this study.

Patterns of spatial data use are changing rapidly, and are expected to continue to do so in the future. Some application areas are relatively mature, with widespread use of spatial data, but in other areas, such as insurance-related applications, use of spatial data and GIS is only just beginning. The same general observation is true of geographic areas—some states and local governments have a long history of GIS use, while others are relative newcomers.

In this situation, there was little to be gained by attempting to make statements about the United States as a whole, or the average GIS user, or the average potential user of GIS. An average forester may be familiar with GIS, but an average insurance broker may never have heard of it. The meaning and significance of "potential use" and "actual use" are thus specific to each sector of the economy, and there seems little point in trying to establish uniform meanings across such widely divergent contexts. These issues amount to problems in defining a population to be sampled, and thus in making inferences from a sample to such a population. We argue that there exists no effective definition of such a population. Instead, we have allowed accessibility and expedience to define the sampling frame for this study, rather than any coherent notion of a population.

A list of 3222 potential respondents was obtained from many sources. Persons who use spatial data in a GIS context and who feel that this is an important part of defining themselves as professionals, often decide to belong to one or more professional organizations that are identified with GIS and spatial data use. In the first category are persons listed in the *GIS in Business '93 Conference Proceedings (GIS World* 1993), persons listed in the *Proceedings of AM / FM International* (AM / FM 1993, 1992, 1991), persons listed on the *Proceedings of GIS / LIS* (ACSM*et al.* 1993, 1992, 1991), and persons listed in the URISA membership directory (URISA 1993). We call this category of respondents the self-selected category. In other cases, there are professional groups that exist to serve professionals who use spatial data in a GIS context; these groups identify professionals to target research and development of GIS. In the second category we developed our list of potential respondents from the State Geographic Information Activities Compendium (Warnecke 1992) and from the International GIS Sourcebook (Parker 1991). These organizations are not the only professional organizations to which professionals who use spatial data and GIS choose to belong.

We chose the groups in the first category because we believed they attracted professionals with a wide variety of application areas, in contrast with other organizations which reflected more narrow application areas. For the same reasons we selected second category sources as organizations that identify spatial data users from a wide variety of application areas. Our sample is from individuals in these two groups. The result, we believe, gave us a list of potential respondents that reflected individuals with a professional interest in the characteristics and use of spatial data without bias with respect to any particular application area. Approximately 60 percent of the list were drawn from the self-selected category and approximately 40 percent were drawn from the third-party identified category.

There was a presumption that the sample population had some minimal technical knowledge of the geographic data they used. This knowledge would be gained through the use of the data within geographic information systems themselves or through the use of products generated from GIS. As shown in the results, a significant portion of our sample population did not fall into either of these categories, but were rather concerned with the policy and organizational aspects of GIS. It is also likely that a large segment of the private sector GIS community was not included in the sampling process simply because of the difficulty of identifying such users due to reasons of business confidentiality or privacy concerns.

Each respondent was identified by state, and by a best estimate of occupation and sampling category. Systematic samplings were then made within each state, occupation, and employment sector to obtain a sample. While the objective of the drawing was to obtain approximately equal samples in each state, each occupation, and each category, the relative lack of potential respondents in certain occupations and states made this impossible. The final sample consisted of 1360 potential respondents.

Stratified random samples are normally used in order to provide statements and inferences of equal reliability in each sector or stratum of a population, when such sectors or strata occur with unequal frequencies. In order to make statements and inferences about the population as a whole from a stratified sample, it is necessary to weight responses within each stratum by a factor reflecting the stratum's sampling intensity. Because there is little to be gained from constructing the average GIS user or average potential GIS user in the United States, we have chosen in this study to report the results of various combinations of subgroups in detail and to not construct a weighting scheme for the entire sample. The entire population of individuals using geographic information is so varied and diffuse that any attempt to arrive at generalized inferences for the entire population would be futile. For this reason we refer to our approach as a "segmented" rather than a "stratified" sampling, reflecting the impossibility of defining a population. We have not attempted to weight responses.

For reasons already expressed, we believe it is easier to define the sampled population within each category, occupation, and state, but that severe differences between these sub-populations exist, particularly across occupations. All results from this study are stated as counts, proportions, or averages for the sample and its sub-samples, and while comparisons are made between proportions, no inferential statements are made regarding any larger population. We have summarized these comparisons in the following pages. However full analyses and data can be obtained from NCGIA.

Although the initial attempt was to collect and analyze responses across each of the 50 states and the District of Columbia, small and possibly heavily biased samples from some states made this impractical. Instead, the states were aggregated into nine geographic regions corresponding to the U.S. Census use of national regions (Census 1993).

Sampling Method

Persons currently using geographic information systems or products generated from GIS were the target of the questionnaire. We focused on people using end products which they knew were produced by GIS as well as "hands-on" GIS technicians and managers. We assumed that the breadth of future users of GIS is already reflected in the current users of GIS, although the proportion of future users in different categories may vary drastically (e.g., there may be a far greater number of business users in proportion to government users in the future, but these business users will be reflected in the current geographic/land information user community).

Sampling was segmented by three methods. First, it was segmented across the government, private, and academic sectors. Second, it was segmented across occupational sectors. Third, it was segmented by geographic area. Priority was given to sampling across a range of occupational sectors.

Occupational sectors were categorized using U.S. Bureau of Labor Statistics bulletin information statistics giving the numbers of people employed in different sectors of the U.S. work place (BLS 1992). A total of 60 occupations which seemed particularly suited for using GIS applications to solve everyday work problems were identified and were aggregated into 30 application areas listed in the questionnaire (see Table 1). Since users might be applying GIS in more than one area, they were allowed to respond with the frequency of GIS use in each of these 30 application areas. Estimates were made on the possible application areas of each member of the sample population on the basis of the potential respondent's affiliation included in his or her address. For example, if a person listed his or her address at a city planning agency, he or she was added to the list of urban and regional planners. These occupational lists of potential respondents were then pared (if needed) and balanced as far as possible across government, private, and academic sectors. Finally, the sample population was balanced as best as possible across geographic regions on a state-by-state basis.

Each copy of the questionnaire was numbered and cross referenced to the name and address to which it was sent. Follow-up letters were sent to those not responding within the first three weeks. Another copy of the questionnaire was mailed to those who did not respond to the second mailing after two weeks.

Questionnaire Content

The questionnaire was 24 pages long and was divided into two portions. The first portion gathered personal information from each respondent, including discipline, GIS applications area(s), employment level, organization background, and GIS experience.

The second portion of the questionnaire asked the respondents to detail their information needs for six classes of geographic information identified by the FGDC as being priority framework data: 1) transportation feature data, 2) water feature data, 3) other well-defined cultural feature data, 4) elevation data, 5) land parcel data, and 6) boundary data. We do not suggest that these are the only possible framework data sets, or that these were necessarily the top priority framework data sets, but wished to limit the choices of data to tie this study in with other work being performed by the FGDC.

A glossary of terms used to define the technical features of framework data was included in the questionnaire. Use of the glossary was encouraged to resolve any ambiguities about the meanings of the terms described.

First, the respondent was asked if he or she used or foresaw a future need to use that class of data in their job. If the respondent answered no, they were requested to skip that class and respond to the next class of data. If the respondent answered yes, they were asked for further information regarding their data needs for that class, including content, tasks for which the data are used, format, geocoding scheme, positional accuracy, vertical accuracy (if needed), updating interval, needs for historical data, and the sources for data currently be-

TABLE 1. Thirty occupational application areas for GIS.

1. Administration	11. Engineering (Civil)	21. Oceanography/Marine
2. Agriculture	12. Engineering (Other)	22. Property/Real estate
3. Architecture	13. Forestry	23. Public relations
4. Banking / Finance	14. Geology/Geophysics	24. Shipping
5. Biology	15. Insurance	25. Social science/Social services
6. Communications	16. Law (except police services)	26. Surveying
7. Construction	17. Legislative	27. Transportation
8. Economics	18. Marketing/Advertising	28. Urban and regional planning
9. Education	19. Medical/Health	29. Utility operations
10. Emergency services (police, fire, etc.)	20. Meteorology/Air quality	30. Wildlife

ing used. These questions are standardized as much as possible for each class, except for the data content questions, which must differ according to the class of geographic data. In all cases, the respondent was allowed to fill in a response if he or she did not find the provided answers adequate.

The survey responses were tabulated and analyzed using Statistical Program for Social Scientists (SPSS) software and custom-developed software. Preliminary results reflect the data needs and uses of the aggregated respondents. Results have also been compiled for various user occupations, for various levels of government and private business, and for various geographic sections of the United States for certain survey questions. Data content, data tasks, data formats, data geocoding schemes, data positional accuracy, and data vertical accuracy were broken down into tables showing the response frequencies for each geographic region, for each application area, and for each employment sector.

Frequency responses were numbered from 1 (never use) to 5 (often use). The application area responses were broken down with separate tables showing the response frequencies across all application area users (those giving a response between 2 and 5 for using a certain application area) and across heavy application area users (those responding with a 4 or 5 for a certain application area use).

Responses in which persons replied that they did not use GIS or products generated from GIS in the performance of their jobs were not processed for technical contents, since such users were asked to not complete the questionnaire. The respondent's geographic location was coded by state based on matching the questionnaire number against the original mailing database address for that respondent.

Questionnaire Results

The first wave of questionnaires was returned by mail during the first week of June, 1994, and collected until mid-August, 1994. An initial 35 questionnaires were returned because of incorrect or out-of-date mailing addresses. A total of 724 valid responses (54.6 percent) were received. Of these, 129 (9.7 percent) responded that they did not use GIS or GIS products. The remaining 595 usable responses (44.9 percent) were keyed into a raw data file and were cross-checked and validated against the range of response values for each question by use of custom software. The number of usable questionnaires was verified by a hand-count of the questionnaires. The numbers of usable responses were tabulated by geographic region, by the discipline of the respondent, and by the application areas that he or she is using GIS or products generated from GIS.

Respondent's Background

Respondent's background was obtained through the questions in Section 1 of the questionnaire. Respondents were asked questions regarding the area of application for GIS, their level within their organization, the type of organization they were with, their total experience using geospatial data, their experience using digital geospatial data, and the amount of work time that they used such data on their job.

Some respondents replied that they filled out the questionnaire based on their organization's use of GIS, not on their own individual use. This would seem to account for the large number of application areas seen on many responses. Of the 30 application areas listed, an average of 10.6 application area uses per respondent was noted (see Table 2). An average of 4.3 application areas were answered with a value of 4 or 5, indicating frequent or heavy use. Five respondents did not select any of the 30 listed application areas, instead listing their individual primary uses of GIS or products derived from GIS as demographic analysis, emergency management, cartographic mapping, planning, and groundwater/drinking-water supply.

Other respondents wrote in answers for other application areas in which they used GIS or products derived from GIS. A total of 81 write-in responses was received. Some respondents wrote in more than one additional application area. Eighteen of these write-in responses concerned water studies or water management in some form. Ten write-in responses addressed environmental issues. Eight responses addressed property management concerns, while six stated mining interests. Most of the other write in responses were very discipline- or task-specific. For example, one person wrote in "military applications" while another person wrote in "crime analysis."

Local government employees accounted for the largest number of responses, 36.3 percent. Of these, 15 percent worked for county governments, 15.6 percent

TABLE 2. Number of GIS application areas by employment sector

	Number of applications					
	1–5	6–10	11–15	16–20	21–25	26–30
Federal	16	7	5	7	0	2
State	85	44	39	24	16	5
Local	50	38	48	29	42	9
Private	39	25	19	9	7	2
Academic	13	5	4	4	2	0

worked for city governments, and worked for other local government agencies, including regional and tribal government agencies. State government employees accounted for 35.8 percent of the responses, while federal government employees responses totaled 6.2 percent. Those involved in education and research (academia) comprised 4.7 percent of the responses. The remaining 17 percent of the responses were from the private sector, primarily consulting (7.6 percent) and services (7.2 percent) based companies. While a larger number of responses from private companies, particularly landholding companies, would have been desirable, these companies were difficult to identify and include in the survey. Our results probably do not adequately represent the uses and needs of such companies.

The respondents held primarily professional (46.1 percent) or middle-management (29.2 percent) positions within their companies or agencies. Fewer classified themselves as holding either upper-management (15.8 percent) or technical (8.7 percent) positions. No one with a clerical position responded to the survey. More than half the respondents had used geospatial data in some form for over 10 years. Roughly a fifth of the respondents had used geospatial data for five years or less. However, more than half the respondents had used digital geospatial data or products for five years or less, while roughly a fifth had used digital geospatial data for more than 10 years.

Data Use and Need Responses

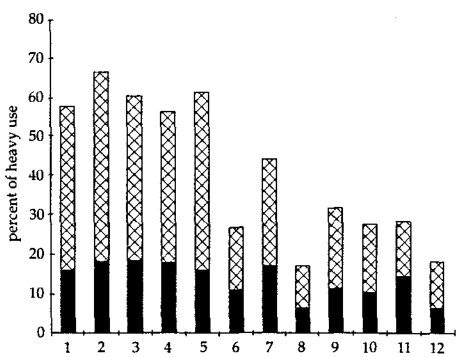
Section 2 of the questionnaire showed that transportation feature data were used or needed by 91.9 percent of the respondents while 85.3 percent used or needed water-feature data, 83.3 percent used or needed other well-defined cultural-feature data, 70.6 percent used or needed elevation data, 80.9 percent used or needed land-parcel data, and 94.3 percent used or needed boundary data. In each of the data categories, people responded that they felt they needed better data by a three-to-one margin.

Data Content Responses

More than half of all the respondents claimed a heavy use or need for transportation data about freeways, highways, local trunk roads, and city or town streets. Slightly less than half need transportation data about railroads. One third or fewer had a heavy need for other transportation data. Responses are shown in Figure 1.

Response rates tabulated for transportation feature data contents across geographic region, all occupational users, heavy occupational users, and employment show no significant deviation from the aggregate responses.

FIGURE 1. Transportation data content responses



1) Fwys 2) Hwys 3) Local 4) Rural 5) Streets 6) Trails 7) RRs 8) Light rail 9) Water 10) Airports 11) Pipes 12) Buses

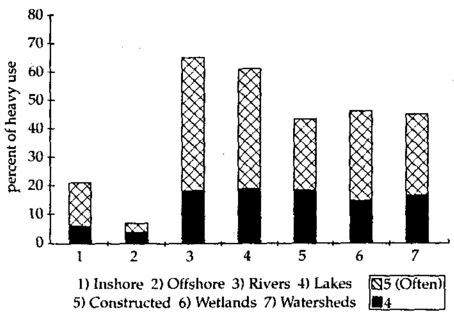
Popular write-in responses included recreational paths (n = 10) and transmission lines (n = 7).

More than half of the respondents noted a heavy use or need for water feature data about rivers/streams and lakes/ponds. Between 40 percent and 50 percent of respondents had a heavy use or need for water feature data about constructed waterways, wetlands, and water sheds. Approximately one-fifth had a heavy need or use for inshore ocean data while less than one-tenth had a heavy need for offshore ocean data.

As might be expected, those respondents living in a region near coastal waters did have a higher need for inshore and offshore ocean data than other regions. Also, as expected, those listing oceanography/marine as a GIS application area had significantly higher needs for ocean data. However, these responses seem to indicate a need for inshore, rather than offshore, ocean data. No other deviations in responses seem significant. Popular write-in responses for additional water feature content were flood plains (n = 6) and water wells (n = 4). Water feature data content responses are shown in Figure 2.

Although more than 80 percent of the respondents replied that they used or needed data about other well-defined cultural features, fewer than 40 percent responded with a heavy use or need for any one particular feature. Responses for all of the listed cultural features were between 20 percent and 40 percent for heavy users (see Figure 3). Deviations from the aggregate response rates were primarily application area oriented. For example, those listing medical/health appli-

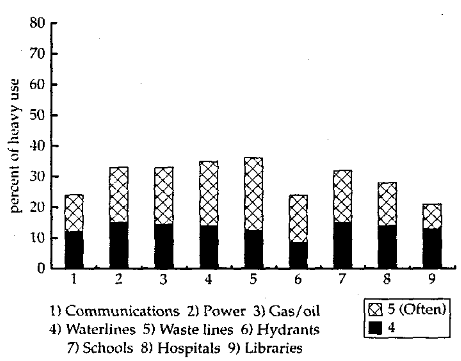
FIGURE 2. Water feature data content responses



cations had a higher than average need for hospital location data. Response frequencies across employment sectors did show a higher need for transmission line data in the private sector than in government and academic sectors. Popular write-in responses for other well-defined cultural feature data included parks or recreational sites (n = 20), commercial structures (n = 9), residential structures (n = 7), and all structures (n = 7).

Digital contours were the most popular content for elevation data with about 40 percent claiming a heavy use or need for digital contours. Federal government employees were more frequent users of digital elevation models (DEMs) than the average user, but no other significant deviations are noted. The only write-in response for additional elevation data contents was for building heights. Figure 4 shows elevation-data content responses.

FIGURE 3. Other well-defined cultural feature data content responses



All of the land-parcel data content items were very popular, with each item receiving a high frequency of use or need among aggregate respondents. More than half of the respondents had a heavy use or need for individual land parcels and public right-of-way data. Consolidated land ownership data public easement data, and private easement data were all cited by more than 40 percent of the heavy users. As should be expected, all land parcel data are most popular among those in property/real estate applications. They are also more frequently noted among administrators, architects, construction users, civil engineers, law applications, and surveyors—and especially among those classified as heavy users in those applications. Predominant use of land parcel data is in local government, followed by private industry. Only two write-in responses were received for additional land parcel data contents; one was for land use, the other for building set-back lines.

City and county boundaries were the most frequently used or needed boundary content items among the aggregate users and each received a response rate of about 70 percent from heavy users (see Figure 6). Other political jurisdiction boundaries were also popular with more than half the respondents citing a heavy use or need for data about states boundaries and other administrative boundaries. Census zone data was heavily used or needed by slightly fewer than half of the respondents. About one-third of the respondents had a heavy use or need for zoning boundary data. Fewer than one-fourth had a heavy need for ZIP code boundary data. Most of the write-in responses appear to be to related to some special political boundary need, such as maintenance or service districts.

FIGURE 4. Elevation data content responses

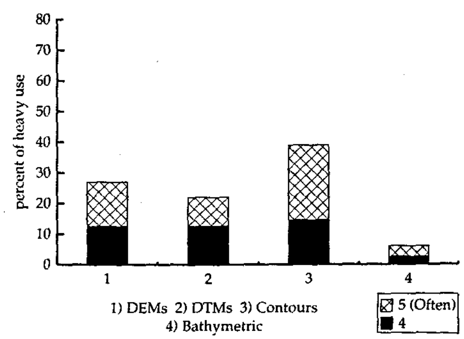
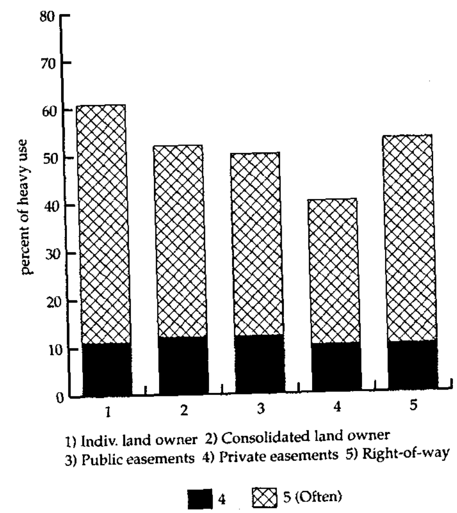


FIGURE 5. Land parcel data content responses



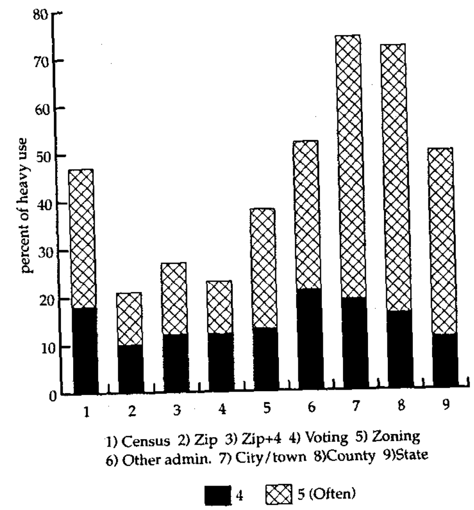
Stratified responses for boundary data contents show that census tract and ZIP code boundaries are very popular among banking/finance, economics, insurance, legislative, marketing/advertising, medical/health, and social science/social services applications. Popular write-in responses include national boundaries (n = 7), tribal/reservation boundaries (n = 5), congressional districts (n = 4), land use areas (n = 4), and traffic analysis zones (n = 4).

Data Tasks Responses

Address matching, where applicable, and site analysis were the most popular data task responses for all data categories. Although address matching was popular, it was much less frequently cited by users of natural science applications—agriculture, biology, forestry, geology, oceanography, and wildlife. It was also much less frequently cited among federal, state, and academic sector employees. Environmental monitoring, however, was significantly more popular among federal and state employees and among natural science applications. Environmental monitoring was the most frequently cited task for water-feature data.

Site analysis and inventorying were moderately popular data tasks. Vehicle routing was a task infrequently cited when offered as a choice. Figure 7 shows an example of responses using transportation-feature data.

FIGURE 6. Boundary data content responses



Data Format Responses

By a large margin, respondents used or preferred data in vector format. Over 60 percent of the respondents in each of the five data categories where vector-format data was offered as a choice cited a heavy use or need for vector-format data. Digital photography, including digital orthophotos, was the next most popular format with around 30 percent of heavy users responding. Imagery and other raster-based data were heavily used or needed by fewer than 20 percent of the respondents in each data category. Digital contours were the most-used or favored format for elevation data with about 40 percent of heavy users responding. Grid and TIN elevation data formats had about 25 percent each of respondents with a heavy need or use. No significant deviations from aggregate responses were noted among the stratified results. Responses for transportation-feature data formats are pictured in Figure 8.

FIGURE 7. Transportation feature data tasks

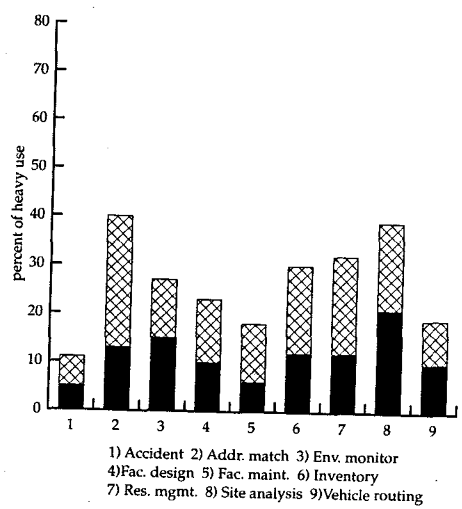
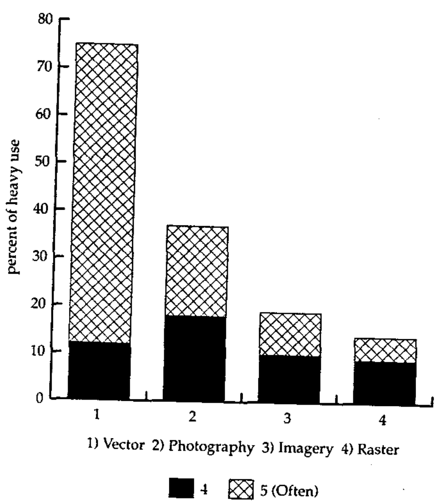


FIGURE 8. Transportation feature data formats



Data Geocoding Scheme Responses

State plane coordinates were highly popular as the geocoding unit for each of the six data categories. More than 50 percent of the respondents in each data categories cited a heavy use of state plane coordinate except elevation data, where about 40 percent of the respondents cited a heavy use or need. Where offered as a geocoding alternative, street addresses were also highly popular. About 50 percent of the respondents cited a heavy use or need for transportation data geocoded using street addresses while more than 40 percent cited a heavy use or need for land-parcel street addresses and about one-third cited a heavy use or need for cultural feature data street addresses. Land-parcel identification numbers were cited as heavily used or needed for land-parcel data by more than 40 percent of the respondents. Latitude and longitude values were used or preferred in all categories over UTM coordinates.

Stratified responses for geocoding schemes suggest that those application area users who do not perform much address matching do not cite street addresses as a frequent need. As noted above, these seem to fall into the area of natural science users—agriculture, biology, geology, meteorology, oceanography, and wildlife. Street addresses are highly cited among all categories by banking/finance, economics, emergency services, insurance, law, legislative, marketing/advertising, medical/health, property/real estate, public relations, shipping, and utility operations application users. Street addresses are also much more frequently cited among local government and private industry employees. Regionally, the U.S. Public Lands Survey System (USPLSS) was not popular in areas where it is not used, such as New England and the Middle Atlantic states. Federal government employees appear to be the most frequent users of latitude and longitude geocodes.

Data Positional Accuracy Responses

Positional-accuracy preferences were primarily in the 1-meter to 20-meter range for all data categories. Many respondents seemed to have trouble answering this question. Several respondents left this section blank, while a few other respondents marked each choice as "never" used with no write-in or other response to the question. Almost 40 percent of the respondents cited a heavy use or need for 1-meter positional accuracy for land parcel data while less than 30 percent cited 1-meter as the most popular heavily used or needed positional accuracy for elevation data. Boundary data and other well-defined cultural-feature data showed heavy use or need frequencies most popular at both the 1-meter and 10-meter positional accuracy. Transportation-feature data and water feature data were most popular at a 10-meter positional accuracy among the respondents citing a heavy use or need.

Responses stratified across application areas show a high response rate among architectural applications for 0.1-meter positional accuracy data across all data types. Construction, emergency services, civil engineering, other engineering, insurance, property/real estate, surveying, and utility operations applications responded with 1-meter accuracy most frequently. The natural science applications—agriculture, biology, forestry, geology, meteorology/air quality, oceanography/marine, and wildlife—seem to accept data with 10-meter positional accuracy. Most other applications seem fairly evenly balanced in their responses to both 1-meter and 10-meter positional accuracy data. One-meter accuracy data is also more popular among local government employees than among other employment sectors. Ten-meter positional accuracy data seems to be the most popular need at state government, private sector, and academic levels.

Write-in responses show some of the user confusion over positional accuracy. One respondent wrote in requirements as meeting the National Map Accuracy Standard. Another respondent required positional accuracy "the same as in TIGER." Other write-in responses along this line include "whatever they give us," "not sure," "as accurate as possible," and "the more accurate the better." Those responding with write-in values gave answers ranging from 1 foot to 30 meters as their required accuracy.

Data Vertical (Elevation) Accuracy Responses

As with positional accuracy, many respondents seemed a little confused about their use or need for vertical accuracy. While nearly 30 percent of respondents cited a heavy use or need for elevation data with 1-meter vertical accuracy, only about 20 percent or fewer did so in the other data categories. The responses were nearly even for and against the need for elevations for transportation-feature data, water-feature data, and other well-defined cultural-feature data. Since vertical information is inherently a part of elevation data, respondents were simply asked to list their vertical accuracy requirements. Only those who used or needed elevation data for transportation-feature data, water-feature data, and other well-defined cultural-feature data were asked for their vertical accuracy uses or needs.

The most frequent vertical accuracy requirements among aggregate users were slightly higher than the positional accuracy requirements, ranging from 0.1 meter to 10 meters. Further investigation may be warranted to examine a possible correlation between the need for elevation information and the accuracy requirements of those who need elevation information.

Stratified response rates for vertical accuracy varied across application uses. As with positional accuracy, architectural applications required the most accurate data, with popular values ranging from 0.01 meter to 0.1 meter. Banking/finance applications also seem to fall within this interval. Construction, civil engineering, insurance, and surveying applications seem to fall into the range between 0.01-meter and 1-meter vertical accuracy. Communications, economics, emergency services, non civil (other) engineering, and utility services applications seem to fall into the 1-meter vertical accuracy range. The natural science applications (agriculture, biology, forestry, geology, meteorology/air quality, oceanography/marine, and wildlife) seem to fall in the 1-meter to 10-meter range.

Data Updating Interval Responses

The respondents did not prefer to have data updated on a periodic basis (e.g., weekly, monthly, or yearly). Land-parcel data and boundary data were cited most fre-

FIGURE 9. Transportation feature data geocoding schemes

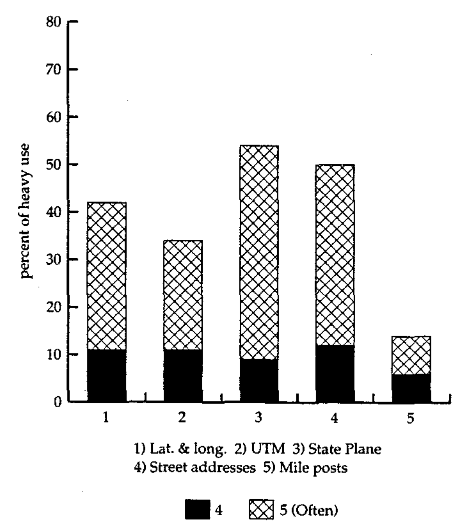
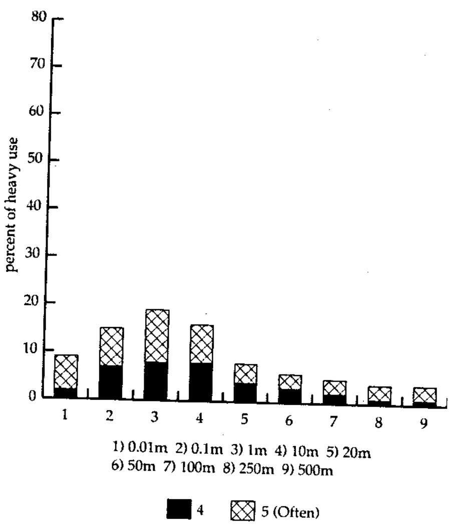


FIGURE 10. **Transportation feature data vertical accuracy**



quently as needing updating as any changes occurred. Several write-in responses were received, but have not all been analyzed to date. Among the write-in answers, intervals of several years are often cited as an appropriate updating interval. However, some write-in responses were to the point, citing "as funds available" and "as funded" as their actual updating intervals.

Historical Data Needs

The respondents' need for historical data varied from category to category. Historical land-parcel data was needed by a margin of about 2 to 1. Historical boundary data were also needed more often than not. The need for historical transportation-feature data, historical water-feature data and other well-defined cultural-feature data was nearly even. Historical elevation data was not needed by a margin of about 2 to 1.

Data Sources

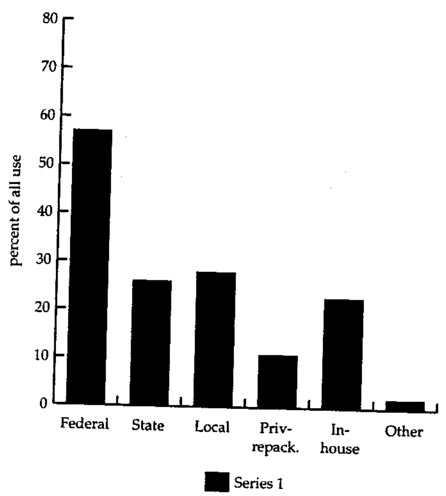
The federal government was the source of most of the data used by the respondents, with the exception of land-parcel data, which was obtained primarily from local overnment sources. State government sources were used to a lesser extent, while almost as many collected data in-house or through private data-collecting sources. Because users could cite more than one source for their currently used data, many did so. Popular sources for federal data were Tiger and USGS quadrangle sheet data, although it was unclear in most cases where users were citing USGS Digital Line Graphic (DLG) data or were digitizing the data from hard-copy USGS quadrangle sheets. The totals for data sources may accordingly add up to more than 100 percent for any of the six data categories. Responses for transportation-feature data sources are shown in Figure 11.

Summary and Conclusions

Summary

The questionnaire respondents were predominantly professional or middle-management persons. Approximately one-third worked for local government while another third worked for state government. Of the remaining third, most worked for private companies. More than half of the respondents had ten or more years experience working with any type of geospatial data, however, more than half of the respondents had less than five years experience working with digital geospatial data. The respondents ranged roughly evenly from casual user of GIS to heavy users of GIS. The respondents were spread across the United States and averaged more than ten different occupational uses of GIS with heavy frequency.

FIGURE 11. **Transportation feature data sources**



Boundary data was the most frequently cited framework data category with more than 94 percent of the respondents using or needing such data. The respondents primarily wanted boundary data for towns, cities, counties, and states in vector format, geocoded using state plane coordinates at a positional accuracy of 1 to 10 meters. These data are used primarily for address matching, demographic analysis, and site analysis.

Transportation-feature data was the second most frequently cited data category with almost 92 percent of respondents using or needing such data. These respondents primarily wanted transportation data about freeways, highways, local trunk roads, rural roads, and city or town streets. They wanted these data in vector format, in state plane coordinates and with street address information, primarily at a positional accuracy of 1 to 20 meters. Only about half of the respondents needed elevation information about transportation data, primarily at a vertical accuracy of 1 to 10 meters. These data are used primarily for address matching and for site analysis.

Water-feature data was the next most popular data category with about 85 percent of respondents using or needing water data primarily about rivers, streams, lakes, ponds, constructed waterways, wetlands, and water sheds. This data is used primarily for site analysis, inventorying, and environmental monitoring. The users primarily want the data in vector format, geocoded using state plane coordinates, at a positional accuracy of 10 meters.

Other well-defined cultural data were used or needed by over 83 percent of the respondents. These data are needed primarily for site analysis and address matching. They are needed in vector format, geocoded primarily using state plane coordinates at a positional accuracy of 1 to 10 meters.

Land-parcel data was cited as being needed or used by over 80 percent of the respondents who wanted land parcel data about single and consolidated land ownership, easements, and right-of-ways. They wanted this data in vector format, primarily geocoded in state plane coordinates, parcel IDs, and street addresses at a positional accuracy of 1 meter. Land-parcel data is used primarily for site analysis, address matching, and inventorying.

Elevation data was used or needed by over 70 percent of the respondents who wanted primarily digital contour data geocoded using state plane coordinates at 1- to 10-meter positional and 1-meter vertical accuracy for site analysis.

The respondents wanted data updated primarily as changes occurred and were split in most cases on their need for historical data. The primary source for digital geospatial data was from the federal government except for land-parcel data where the primary source was local government.

Conclusions

Many current GIS users appear to be using GIS across more than one disciplinary application area. An average number of almost ten applications were cited per user. This could mean that many of the problems being solved through GIS are interdisciplinary in nature. It could also signify that many GIS tasks are being performed by central facilities on behalf of a variety of departments. It might also reflect user responses to organizational uses, rather than individual uses, of GIS. Since more than one application area was cited by many users, possible correlation in the answers stratified by occupational area use will need to be further investigated.

Many current GIS users appear to carry a map analogy into their technical knowledge of GIS. While maps have a limited use based on a given scale, GIS databases have a much wider range of functionality because of their ability to display data at many scales and their ability to perform routine analytical functions with little or no human intervention. This apparent lack of technical knowledge about GIS among many current users also becomes apparent when viewing the response rates for elevation data formats. Digital contours, an analogy carried directly from maps into GIS, are highly popular while grids and triangulated irregular networks (TINs), data formats with which most GIS databases address elevation information, are much less popular.

Consequently, technical issues such as data positional and vertical accuracy become of much greater concern in GIS databases (Goodchild and Gopal 1989). However, many respondents had difficulty in addressing such issues. Several possibilities arise that might explain this phenomenon. First, users might not have documentation (metadata) that can readily give them answers to such technical questions. Second, users may depend on secondary positional and vertical accuracy measures, such as map scale and contour interval, to provide accuracy measures. Third, users might depend on the judgment of others to provide them with data of sufficient accuracy for their individual needs. Fourth, users might be using whatever data sets they find available without worrying about the possible technical qualities of that data. Although much has been done to promote concerns about these issues—such as the FGDC Metadata Content Standard and the recent International Symposium on the Spatial Accuracy of Natural Resource Data

Bases held in Williamsburg, Virginia, in May 1994—perhaps more education and emphasis on understanding the implications of these issues may be in order.

Differences in user requirements appear to be primarily application oriented. Some regional differences were noted, primarily concerning proximity to oceans and the use of the U.S. Public Lands Survey System. Differences among employment sectors seemed to be primarily in areas related to the amount or scale of data being used. This is particularly apparent in the geocoding scheme responses, where federal government employees, users of typically large spatial areas of analysis, seem to prefer a scheme that will deal effectively with the representation problems of those areas—namely latitude and longitude coordinates. Street addresses, primarily a geocode useful for relatively small spatial areas of analysis, are not very popular among federal employees. However, street addresses, like state plane coordinates, are much more popular among users who typically deal with relatively small spatial areas of analysis, such as local government and private sector employees.

Future analyses planned for this data include a look at the possible clustering of GIS technical requirements that combine distinct regional, occupational, and employment sector needs. As noted in the section relating the positional accuracy responses, it appears that certain GIS applications appear to have similar data technical needs with other applications. The question arises whether GIS users could naturally fall into a limited number of user categories.

We have attempted to find the current digital geospatial data technical requirements of users with this study. We do not attempt to predict future needs. While some technical requirements, such as data format and data contents, may remain stable for the foreseeable future, other technical requirements, such as positional and vertical accuracy, may increase in importance with time. This study attempts to define technical guidelines for selecting current digital geospatial data sets that may be useful for the greatest numbers of users which may then be prioritized for inclusion in the National Spatial Data Infrastructure.

This study reflects perhaps the first large-scale effort to sample the U.S. population of GIS users in a systematic fashion, regarding their current and future uses of geospatial data and related technologies. The analyses reported here are a small fraction of the results which might be extracted from the data, and we anticipate that others may have specific questions they wish to ask of the database. NCGIA will provide copies of the database on request to any bona fide group or agency wishing to conduct further analyses.

Acknowledgements

This research was conducted for the Federal Geographic Data Committee with funding provided by the U.S. Geological Survey under project agreement number LA#93-5940-5033. We thank Nancy Tosta and Mike Domaratz of the National Mapping Division, USGS, for their interest in this research.

References

ACSM, ASPRS, AAG, URISA, and AM/FM. 1993, 1992, 1991. Proceedings of GIS/LIS.

AM/FM (1993, 1992, 1991). Proceedings of AM/FM International. Aurora, CO: AM/FM International.

Bureau of Labor Statistics. 1992. *Outlook: 1990–2005: Occupational Employment* Washington, D.C.: The United States Bureau of Labor Statistics.

Budic, Zorica D. 1993. "GIS Use Among Southeastern Local Governments." *URISA Journal*, Vol. 5(1): 4–17.

GIS World. 1993. In *GIS in Business '93 Conference Proceedings*.

Goodchild, M.F. and S. Gopal. 1989. *Accuracy of Spatial Databases*. London: Taylor and Francis.

National Research Council. 1994. *Toward a Coordinated Spatial Data Infrastructure for the Nation*. Mapping Science Committee, National Research Council. Washington, D.C.: National Academy Press.

Parker, H. Dennison. 1991. International GIS Sourcebook 1991–1992. Ft. Collins, CO: GIS World.

Sandberg, Brenda Caroline.1992. *A Geographic Information Systems User Survey: The Emerging GIS Community in Michigan*. Center for Remote Sensing, Michigan State University.

URISA. 1993. URISA Membership Directory 1993–1994. Washington, D.C.: Urban and Regional Information Systems Association.

Ventura, Stephen J. 1991. "Implementation of Land Information Systems in Local Government—Steps Toward Land Records Modernization in Wisconsin." Madison: Wisconsin State Geographer's Office.

Warnecke, Lisa. 1992. State Geographic Information Activities Compendium.