

**Exploratory Studies of the  
NSGIC/FGDC Framework Survey:  
LOOKING AT THE STATE OF THE NATION**

Final Report To

The Universities Consortium for Geographic Information Sciences  
And  
The Federal Geographic Data Committee

Principal Investigator: Dr. David L. Tulloch  
Grant F. Walton Center for Remote Sensing and Spatial Analysis  
14 College Farm Road  
Cook College -- Rutgers University  
New Brunswick, NJ 08901-8551

December 28, 2000

## **REPORT OUTLINE**

### **COMPLETION OF UCGIS GRANT**

?? **Introduction**

### **A NATIONAL FRAMEWORK SURVEY**

?? **NSDI and Framework**

?? **History of the Framework Data Survey**

?? **Overall Goals of the Survey**

?? **What is the Framework?**

?? **Framework Data** – Seven themes of geographic data

### **UNDERSTANDING THE SURVEY RESULTS**

?? **Development of the Survey Instrument**

?? **State Coordinators**

?? **Response Rate Issues**

?? **Sampling**

?? **State Sampling Frames**

### **NATIONWIDE COUNTY-LEVEL RESULTS**

?? **County Level Analysis**

?? **Data Dissemination**

### **TRANSPORTATION FRAMEWORK DATA FINDINGS**

?? **Transportation Framework Data Responses**

?? **Transportation Data Quality**

?? **Transportation Metadata**

?? **Transportation Data Investment**

### **FUTURE APPLICATIONS OF THE FRAMEWORK SURVEY**

?? **Creation of Data Producer/User Phone Book**

?? **Follow-up Survey**

?? **Conclusions**

### **REFERENCES**

**APPENDIX A:** County-level Production of Framework Data: Pieces of a National Spatial Data Infrastructure?

**APPENDIX B:** Looking Back and Looking Ahead: The State of Framework Data

**APPENDIX C:** National Survey Results Regarding Production of Framework Transportation Data

## COMPLETION OF UCGIS GRANT

This document is the final report for the project entitled, “Exploratory Studies of the NSGIC/FGDC Framework Survey: Looking At The State Of The Nation.” It has been prepared primarily for submission to the Federal Geographic Data Committee (FGDC) and the Universities Consortium for Geographic Information Sciences (UCGIS) by Rutgers, the State University of New Jersey. The report signifies the formal completion of the project as granted by the UCGIS. The project proposal (Submitted in June 1999) suggested four deliverables:

- ? ? Meeting with FGDC Staff;
- ? ? Presentation of results;
- ? ? Written findings submitted to a refereed journal; and
- ? ? Written report of findings.

Although the meeting with the FGDC staff was never made in Reston, but a high level of contact was maintained between the project participants and the FGDC staff. This includes not only meetings at national conferences (including the UCGIS 2000 Annual Assembly) but also significant email and phone interactions.

The second deliverable was satisfied by the presentation of a paper entitled, “Looking Back and Looking Ahead: The State of Framework Data” at the 2000 Annual Conference of the Urban and Regional Information Association (URISA). This presentation was also accompanied by a paper of the same title in the conference proceedings (Tulloch 2000).

The third deliverable was that findings be developed into a manuscript and submitted to a refereed journal. That was fulfilled in October 2000 with a manuscript submitted to the URISA Journal. That manuscript, “Exploring County-level Production of Framework Data: Analysis of the National Framework Data Survey” (Tulloch and Fuld 2000), is available at the URISA web site throughout the peer review process ([http://www.urisa.org/Journal/Under\\_Review/articles\\_under\\_review.htm](http://www.urisa.org/Journal/Under_Review/articles_under_review.htm)). It is also included as Appendix A to this report.

The final deliverable is this written report. It includes information intended to guide future analysis of the data by other interested parties, and describes both research activities and research findings. However, the research findings presented herein are not anticipated to be the final reporting of survey analysis. This reflects that the authors expect to continue exploring the data and writing about the survey for additional publications.

## **A NATIONAL FRAMEWORK SURVEY**

### **NSDI and Framework:**

In 1993, after more than two decades of attempts to coordinate or consolidate both national and federal geospatial data efforts, a new concept emerged known as the National Spatial Data Infrastructure (NSDI). Shortly thereafter, the FGDC was created to champion this vision and strive to make it a reality. The NSDI is intended to coordinate the production and dissemination of surveying and mapping data across the United States. Today, as the nation seems to be actively moving away from traditional analog cartographic products towards digital mapping, the NSDI is recognized as an important way to serve the vast public with data from a very wide variety of sources. Before the NSDI can come to fruition, there are several important issues that must be addressed to allow for straightforward guidelines for the searching, transportation, integration, and update of these data.

In order to achieve the goal of a single NSDI, the FGDC has identified the five areas in which work needs to be advanced: clearinghouse, standards, metadata, partnerships, and framework. While the FGDC is guiding the development of many of these elements, it is avoiding a traditional top-down approach. Instead, a recent focus has been placed on investigating a new organizational structure – a chaordic alliance. As of the writing of this document, the GeoData Alliance has been formed to explore more creative structures for advancing the NSDI concept with updates posted at <http://www.fgdc.gov/GeoAll/index.htm>.

### **History of the Framework Data Survey:**

The Framework Data Survey was initiated in 1996 through a cooperative agreement between Federal Geographic Data Committee (FGDC) and the National States Geographic Information Council (NSGIC). The NSGIC is a national organization of state GIS coordinators, and state government geo-spatial data producers and users. FGDC is a committee of federal organizations working with states, local governments, and private organizations to share geographic data through the NSDI. Both organizations hold an interest in facilitating greater data sharing and access.

A major emphasis of the Federal Geographic Data Committee, in attempting to assemble the National Spatial Data Infrastructure, has been the development of the concept of Framework. “The framework forms the data backbone of the NSDI. It has three aspects; data, procedures and technology for building and using the data, and institutional relationships and business practices that support the environment. The framework is designed to facilitate production and use of geographic data, to reduce operating costs, and to improve service and decision making” (Tulloch 2000).

To better understand the nature and status of Framework Data throughout the nation, the FGDC awarded NSGIC a \$200,000 grant to conduct a national survey. While the NSGIC worked with FGDC to develop a core grant management team, it also held several large meetings throughout the survey process to foster an open and inclusive process of survey construction as possible.

These meetings included a preliminary planning meeting held in Chicago, Illinois in the fall of 1996 where the goals of the survey were discussed among more than 30 participants. The final consensus about the goals of the survey and several initial categories of questions were tentatively agreed upon. A second meeting was arranged at the NSGIC mid-year meeting in Chicago in spring 1997 at which the group present edited a preliminary survey instrument with

more than 50 people participating in the decisions about the survey questions. The survey was finalized by a smaller meeting of the grant management team in August 1997 in Boston, during which time final distribution plans were agreed upon. With the distribution process of the survey underway, subsequent meetings of the management team (including meetings in Washington, DC, Reston, VA, and Annapolis, MD) were supplemented with weekly conference calls to work through various issues involved in the distribution of more than 15,000 surveys.

### **Overall Goals of the Survey:**

The overall main goal of the survey was to measure the progress of the nation's framework activities in various organizations, including state, county, academic and private level organizations. In addition, due to the complexity of the survey coordinators' distribution decisions (sampling designs), a goal was created to include participation by all 50 states and to capture data for all counties within all 50 states (Sommers 1999).

Future goals by the FGDC and NSGIC include plans to construct the National Spatial Data Infrastructure. The FGDC is working towards developing a National Spatial Data Infrastructure (NSDI), which would be a source of the best available, publicly accessible geospatial data for users across the country. Since the highest quality data are often locally or regionally produced, the FGDC is interested in identifying local and regional sources of data to populate the NSDI. Specifically, the FGDC has identified 7 Framework Themes that they consider the immediate priority for creating the NSDI. The data collected in the survey are to be part of a national assessment of the currently available data on framework activity in order to shed light on the possibility of creating a national database of framework data. The NSDI will provide the opportunity for all levels of organizations from all 50 states to share data. Overall, the NSDI is a national effort coordinated by the FGDC that responds to a critical need: to improve the ability to create, use, find, and share geographic data (Sommers 1999).

The NSDI is intended to provide a consistent reliable means to share geographic data. Its goals are cost savings for geospatial data collection, enhanced use of geospatial data, and better decision-making using geographic information. The NSDI encompasses the technologies, politics and people necessary to promote geospatial data sharing throughout all levels of government, the private and non-profit sectors, and the academic community. Also includes the technology, policies, standards, and human resources necessary to acquire, process, store, distribute, and improve utilization of geospatial data. The NSDI is an umbrella under which organizations and technology interact to foster activities for using, managing, and producing geographic data. In September 1996, the FGDC and the NSGIC entered into a cooperative agreement to research: Building the NSDI: Measuring progress of Framework activities." Four areas of NSDI activity include: the Clearinghouse, Standards, Partnerships, and Framework Data (Tulloch and Robinson 2000).

The survey is a collaborative effort between the Federal Geographic Data Committee (FGDC), the National States Geographic Council (NSGIC), and state organizations across the 50 states. The Survey measures the amount, type and quality of geographic data created, updated, integrated and distributed by various organizations. In addition, the FDGC and NSGIC were interested in learning the degree to which organizations might be able to share and their willingness to do so as part of a National Spatial Data Infrastructure (NSDI). The overall goals of the survey were: to assess existing coverage of framework data, to estimate financial investments being made; and to examine institutional activities currently taking place (Tulloch and Robinson 2000). An additional object of the survey was to provide a snapshot of framework

data, particularly as a baseline for future longitudinal research. The survey identifies gaps and overlaps in framework coverage and supports efforts towards improving framework data development activities (Tulloch and Robinson 2000).

### **What is the Framework?:**

The NSDI framework is an initiative to develop a readily available set of basic access to these data. It includes the information, operational environment and technology to provide access to these data, and the institutional setting to sustain its development. Framework data are intended to make the creation of other geospatial data sets easier and less costly. Framework data should be: 1) a preferred data source – the best available data; 2) easy to use with other data; 3) accessible at no more than cost of distribution; 4) developed through cooperation with stakeholders, and 5) constructed from the ground up at the local level (Tulloch and Robinson 2000).

### **Framework Data – Seven themes of geographic data.**

- ?? Transportation features used to move people and goods from place to place, including roads, railroads, trails, parks, airports, and waterways and related features such as bridges and tunnels.
- ?? Hydrography – digital data regarding the location and attributes of surface water features, such as streams, rivers, lakes, ponds, canals, and ditches.
- ?? Elevation Data – digital data regarding the vertical distance from a datum to a point or object on the earth's surface.
- ?? Digital Orthoimagery – digitally formatted aerial photography or satellite imagery from which displacements caused by terrain relief and sensor tilt have been removed. The result combines the image characteristics of photograph with the geometric qualities of a map.
- ?? Government (Political) units – boundaries – the geographic extent of units of government, such as states, counties, cities and other incorporated places, town and townships, American Indian and Alaskan native regional boundaries.
- ?? Cadastral Reference System – surveys carried out for the purpose of subdividing land for settlement.
- ?? Cadastral Land Ownership – the geographic extent of the past, current, and future rights and interests in real property, including the spatial data necessary to describe its geographic extent. A cadastre is an official register of the location, quantity, value and ownership of land.

## UNDERSTANDING THE SURVEY RESULTS

### Development of the Survey Instrument:

The survey instrument was developed through a process that reflected as much on the complexity of the cooperative agreement as it did on the nature of the information that was desired. A group of state GIS coordinators, academics, and FGDC employees developed a consensus based survey in digital form that respondents would complete and return. The survey was designed with 118 questions in 3 sections: an initial section collecting basic information about the individual and organization responding to the survey, a section asking general questions about geospatial data within their organizations, and, a section asking specific questions about each framework theme's geospatial data. (Tulloch and Robinson 2000).

Section One: Collects information about all respondents so that they can be contacted in the future about their geospatial data and their survey responses. This section also collected information about the jurisdiction of the respondent's organization.

Section Two: Collects information about the organization and use of geospatial data. Additional questions asked about the GIS software packages used and the incentives that would most likely foster future participation in the NSDI. The first important question to filter out organizations that do not deal with digital geographic data was question 12, where the respondent was asked, "Do you create, update, integrate, and/or distribute digital geographic data?" Those responding No, were then finished with the survey and were to send it back to the FGDC. The next set of questions concerned sharing geographic data, participating in coordinating councils on geographic data, adhering to a policy on geographic data. These questions as well as a question about which part of the respondent's organization deals with geographic data was asked to all those respondents who initially stated that their organization does utilize digital geographic data in some manner.

The next crucial set of questions in the survey falls under the heading Framework Data. In question 23, respondents were asked, "Do you create, update, integrate, or distribute framework data as defined above (defined with the seven themes explained earlier)?" Response categories include only Yes or No. Respondents stating, No, their organization does not deal with framework data were then finished with the survey and were to send it back to FGDC. Only those respondents who answered yes to both question 12, about digital geographic data and to question 23, about framework data, answered the rest of the questions that focused on the seven framework themes. This is very important to bear in mind when interpreting and analyzing the survey findings.

<b>12. Do you create, update, integrate, and/or distribute geographic data?</b>	<b>Raw Number</b>	<b>Percent</b>
<b>Yes</b>	3688	69%
<b>No</b>	1660	31%
<b>N=5348</b>		

<b>23. Do you create, update, integrate, and/or distribute framework data as defined above?</b>	<b>Raw Number</b>	<b>Percent</b>
<b>Yes</b>	2686	50%
<b>No</b>	956	18%
<b>N/A (Includes respondents eliminated by answering “No” to Q 12.)</b>	1706	40%
<b>N=5348</b>		

Section Three: The third and largest section asked specific questions about each of the seven framework themes. Each theme began with a question determining whether the respondent’s organization ‘create, update, integrate or distribute framework data’ for the specific theme. Respondents provided a list of the types of data their organization can provide in each theme (e.g. roads, railroads, waterways, trails, bridges, airports, ports, and tunnels as part of a transportation theme), specifics about the nature of data within that theme (e.g. whether transportation data includes street addresses or linear reference system information), and estimates of completion. In addition, for each theme respondents were asked whether they maintained metadata, the amount of their annual financial investment in the theme, and the accuracy of the data. (Tulloch and Robinson 2000)

The survey distributed to the overall population of 15,000 organizations was standardized. Each respondent/organization received the exact same questionnaire. While there were 50 different state coordinators who handled the survey distribution process, the survey itself was the same for all participants.

The information collected in the survey focuses on those organizations that already create, update, integrate or distribute framework data. Therefore the survey does not include organizations that utilize geographic digital data on non-framework themes, like wetland mapping. Also the survey findings do not include organizations that may want to participate but cannot afford the cost or are in the process of beginning to utilize framework data.

Due to the complexity of surveying organizations in all 50 states, the NSGIC committee determined that a state coordinator for each state should be appointed and then instructed on survey distribution within his/her particular state. The reasoning behind this decision, as opposed to a central survey distribution process through NSGIC/FGDC was intended to allow the identification of more robust mailing lists. Since the NSDI is premised on regional or state coordinators, this survey process was meant as a way of establishing some long-term relationships that might later assist the development of the NSDI. When compared to a national survey distributor, appointing a state coordinator has the possibility of providing more accurate and appropriate mailing lists. The state coordinator approach also provided letters from an individual who was more likely to be familiar to respondents. By emphasizing the state-level benefits, it was hoped that respondents would feel more compelled to respond to the survey.

**State Coordinators:**

First, NSGIC identified a survey coordinator for each state. Initially, coordinators were chosen through NSGIC membership. However, when a NSGIC member was not available or in states where NSGIC/FGDC did not have an obvious choice, state coordinators were identified through other channels. While it is difficult to determine how many states relied upon survey

coordinators who had previously not served in any formal state coordinator capacity, it is important to consider this as a possible distinction underlying variation in state distribution patterns and results. State coordinators who had networks of geographic data users did, in fact, secure more responses than state coordinators not connected to the NSGIC. This should come as no surprise since many of the NSGIC representatives are officials mandated to build networks of contacts. These contacts would know the state coordinator and take seriously a request to complete a survey. With local and state agencies in each state using GIS in different ways, it seemed that only a local person could fully identify likely data producers. For example, in Ohio the survey population needed to include both the county auditors and engineers (two distinct offices). But in Arkansas, while the state coordinator used a database of people that he believed to be using the data, he placed emphasis on county emergency services because they had placed a high priority on automation. In addition, the state coordinators may bring an additional piece of bias to the process. Some coordinators are appointed to coordinate geo-spatial data statewide, but service within a specific department, such as Environmental Protection or Transportation. As a result of their immediate duties (or academic training), these coordinators may provide a biased list of survey recipients, although those recipients may also be more motivated to respond.

The coordinators began distributing surveys in October 1997. In addition to the standardized survey, state coordinators were given uniform instructions as well as advice on how to distribute the survey. Therefore, while each state coordinator had a unique list of organizations, there were some important commonalities in the survey distribution process. First, each state coordinator developed a mailing list of likely framework data creators and distributors within their state. While coordinators were encouraged to include a representative of every county in their state, most did not receive explicit instructions to do so. The state coordinators then assembled and mailed the survey packet to the respondents who when completed and returned the surveys to NSGIC. The final surveys were returned by October 1998. The type and number of organizations sampled varies among states, ranging from counties and state agencies to a wide scope of organizations, including municipalities, regional planning commissions, local offices of federal agencies.

To create uniformity in the process, the State Coordinators were sent initial letters describing the survey and its importance. Next a Cover Letter and survey was sent by state coordinators to selected respondents in each state. The Cover letter stated "In order for us to make a realistic assessment of our statewide status we need every survey returned, even if you are not responsible for any digital spatial data." While 31% of the 5348 respondents stated that their organization does not participate in digital spatial data, there is strong reason to believe that many organizations that do not utilize spatial data did not return the survey and a strong factor in the high non-response rate.

According to the instructions for state coordinators, each state coordinator was to develop its own list of survey recipients. The list was to be as complete a list as possible of all government agencies at all levels that might create, maintain, inventory, distribute, or use digital spatial data. While this design increases the likelihood that those receiving the survey do in fact participate in geospatial data, it creates different sampling frames within each state. The State Coordinators were also instructed to encourage the respondents to feel that their participation is important and beneficial to their state as well as themselves. Incentives for respondents completing and returning the survey included free software, ArcExplorer from ESRI, a book Zeroing In: GIS at Work in Local Government, and a CD-Rom of ESRI Maps & Data. These incentives were provided as gifts from ESRI.

The survey data represent only those organizations that were selected to receive surveys and also that responded. The sample composition varies among states. The survey was not designed to produce a sample of a specific percentage of each organization type. In this sense, the sample may be skewed. Cross-state comparisons cannot be performed reliably on the current data set because the sample composition varies across states (Tulloch and Robinson 2000). The survey instrument distribution process was designed as a way to allow the NSGIC and FGDC to reach as many different possible data sources with their limited resources.

### **Response Rate Issues:**

There are numerous issues concerning the response rate and interpretation of the survey findings. Before discussing some of the survey findings, it is extremely important to understand the limitations of the data.

Overall, a total of 15,000 surveys were sent out to potential respondents, including federal, state, local private and other GIS data producers. Since the state coordinators were not given a specific number of surveys to send out to potential respondents in their states, this does not necessarily mean that each of the 50 state coordinators sent out 300 surveys within their state. This is one important factor to understand, that due to the survey distribution process, there was not necessarily an equal number of surveys sent out within each state. (if a central sampling design had been undertaken, then a standard number of surveys per state would have more likely been distributed). Therefore, before even examining the survey response rates within each state, it is necessary to recognize that due to the state coordinator's familiarity with his/her state, different numbers of surveys were initially sent out within each state. One reason for these differences has to do with the appointment of the state coordinators. State coordinators who sent out more surveys are more likely better connected to the digital geospatial user community within his/her state.

Second, each state coordinator determined which specific organizations should be surveyed within his/her state. The types of organizations were uniform and fall under the categories; federal agencies, regional agencies, state agencies, local agencies, private companies, utilities. More specifically, surveys were sent to federal, state, regional, county, city/town, academia, private and tribal organizations. However, even with this uniformity, the state coordinators based on their own knowledge of GIS users throughout the state, determined the specific organizations that should be sent the survey. In other words, based on factors such as the size of the state, the complexity of the state's governmental organizations, jurisdiction of each organization, the geographic location of the state, etc., and the state coordinator's knowledge and discretion, the number of different organizations initially surveyed varies considerably. What this means is that each state has a sampling frame that can vary by total number and by number of specific organization types. Again, before even looking at the survey findings, it is important to recognize the impact of 50 unique sampling frames on the analysis of the survey findings.

To reiterate, while each state coordinator was instructed on the survey distribution process and collection, there was not uniform sampling design for each state. Since the sampling frame for each state was not determined before instructing the state coordinators on survey distribution, the state coordinators created their own within-state list of potential GIS users. This means therefore, that overall there was not a random sample used within each state and there was not a random sample of each category of organization within each state or across states. Instead we should think of each sampling design, per state, as convenient and purposive. As a result

there are 50 different sampling designs – one for each state. This impacts, therefore, both state comparisons of GIS framework data as well as generalizing to the country-wide level.

In fact, the survey distribution process was originally intended to be a census of all potential GIS users across the United States. The initial plan was to send out surveys to all organizations that might utilize digital spatial data and more specifically framework data and survey all these organizations to find out the state of digital framework data country-wide. However, because each state coordinator determined how many and which organizations to survey, and there were varying response rates within each state, the data do not represent a census or population of framework data users. Instead, these data are limited in generalizability and representativeness of country-wide framework data users, however the data analyzed still paints a rich portrait of various state's framework data users. Data do not necessarily reflect framework activity in a particular state as a whole. All analyses must take into consideration the specific sampling and response biases of a particular state.

One reason why a census was attempted initially in the survey distribution process was to begin to develop a cooperative relationship with framework users within the 50 states. If initially a sample was selected from a list of potential organizations, then the organizations not selected would not have been contacted. While sampling improves the generalizability of the findings to generalize from a smaller group to a larger group, it is not the best method for establishing relationships among GIS users. As discussed later in this report, one of the reasons behind the decisions concerning survey distribution is directly linked to the construction of a 'phone book' of framework data users.

### **Sampling:**

One of the main difficulties in analyzing these data stems from the methods of survey distribution employed. The survey distribution process was very complicated and did not follow the rules of traditional probability sampling. Therefore it is important for readers to fully understand how the surveys were distributed throughout the 50 states and the impact this process has on the generalizability and representativeness of the survey data. In other words, how confident can we be in the accuracy of the data collected to represent the entire population of all geo-spatial digital users throughout the country? This question is significant to the way in which the surveys were distributed and what we can say about the results.

First, when conducting surveys, there are several ways to go about this process depending on the researcher's goals. Sampling is the process of selecting a set number of observations from a larger set of observations called a population, in other words, selecting those people or organizations to study and to distribute the survey to. A population is the theoretically specified aggregation of study elements (Babbie 2000). For the framework survey, the population under consideration that the committee was interested in studying was all organizations that might use geo-spatial digital data throughout the nation (more specifically within each of the 50 states) to measure the status of digital data being used. As stated earlier, the overall goal of the survey was to collect information from organizations within each state on geo-spatial data and the status of that data in order to measure the progress of the nation's framework activities. Initially the survey distribution process was to reach all potential users of geo-spatial digital data within the categories of organizations chosen by the survey committee. Given this goal of finding out about the status of geo-spatial digital data among all potential user organizations, the committee's survey distribution can be thought of as initially using a census. A census is defined as the study of the entire population of interest (Shutt 1996). However, given the process by which the

committee appointed state coordinators who in turn distributed the survey, a census that would have been all 15,000 surveys returned did not occur.

In order to measure accurately the state of the nation's progress of geo-spatial digital data, different types of sampling techniques could have been undertaken. Overall, researchers group sampling techniques into 2 broad categories: probability and non-probability sampling. Again if the entire population is studied then sampling is not needed. A sample is a subset of organizations chosen for the study (Shutt 1996) and a population is the larger collection of elements to which we will generalize our sample findings (Shutt 1996). Specific sampling techniques allow researchers to decide and control the likelihood of specific individuals being selected for study (Babbie 2000). Random (probability) sampling is the method of drawing a portion (sample) of a population so that all possible samples of a fixed size have the same probability of being selected. Probability sampling methods involve the selection of a 'random sample' from a list containing the name of everyone in the population you are interested in studying. This list is called the sampling frame, all the potential elements that might be selected for surveying. Probability sampling is more accurate because it allows researchers to determine the likelihood that the sample selected represents the larger population and based on statistical methods, the results of a sample can be generalized accurately with a certain degree of precision to the larger population. Opinion polls and other large-scale surveys are the best examples of the use of probability sampling to generalize the attitudes of a small group of people to a larger group of people with a certain level of statistical confidence in the results. To provide useful descriptions of the total population, a sample of individuals from a population must contain essentially the same variations that exist in the population (Babbie 2000). When we speak of sampling bias in connection with sampling, this means that those selected are not typical or representative of the larger populations that they have been chosen from.

The other category of sampling is non-random or non-probability sampling where a list of all possible respondents cannot be determined prior to study and therefore statistical probability methods cannot be used to determine the representatives of the sample to any larger population. Therefore, non-probability samples are used when you are studying hard to find populations, expert populations and other groups that are not easily listed in a sampling frame or you do not fully know your total population of interest. While non-random sampling does not provide the statistical accuracy as probability sampling, it is employed for populations that are more difficult to study within probability sampling standards. Overall, decisions about sampling should stem from the research questions under consideration and the population of interest.

For the framework study, each state coordinator was instructed to create a list of all potential users and then to distribute the survey to each of the members on the list. Based on this procedure, a probability sample was not used to study a smaller subset of organizations from a larger population. First, each state coordinator composed his/her own list of organizations and then distributed the surveys to each of the organizations on the composed list. Therefore, the lists within each state were not uniform. In fact, the sampling techniques do not fall neatly into either probability or non-probability sampling. We can think of the sample as most resembling purposive non-probability sampling in which organizations that were potential digital data users were selected for study. In this regard, the results are not easily generalizable to all geo-spatial digital users neither in each state nor across states. This adds a layer of complication to the analysis of the findings.

In addition, because there was a response rate overall of around 35%, there are 65% of surveys not returned. Without knowing exactly who these organizations are, it is difficult to

generalize from the returned surveys to the whole population, the entire 15,000 organizations surveyed. There is both built in bias with a non-probability sample and bias issues with the non-response rate.

Even though there are limitations in our ability to generalize the data to all spatial data users in the United States, there is still a considerable amount of important and useful findings in these data. We are suggesting that those interested in analyzing these data first examine the information provided about each state's particular sampling design (described below). This information is critical in that it clearly indicates how state's differed in the initial survey distribution process and issues related to how best to analyze these data.

### **State Sampling Frames:**

For more detailed information and data about particular state sampling designs, response rates and biases, see [http://www.fgdc.gov/framework/survey\\_results](http://www.fgdc.gov/framework/survey_results). Detailed information on all fifty states is included. For each state's information, the following uniform paragraph is included:

For each state, the target population was all organizations thought likely to be active in framework data. This includes organizations from all levels of government, academia, utilities, non-profit organizations, and private sector firms. The target population excludes all organizations that are not thought likely to be active in framework data. This decision was made by each particular state coordinator.

This information retrospectively attempts to recreate the sampling frame in order to estimate the sample response rate for each organization type. In addition, in many cases, a single organization was sent more than one questionnaire to different departments and a count of this overlap is also included in the response count. Also included is a description of specific details about the data for each state and advice on response and bias issues for those analyzing these data.

Included in the breakdown of information on each state is the following:

1. Table of subgroups by organization type by sample, sample responses, response rate.
2. For specific organization types, a breakdown of response rates by population size to illustrate whether non-response rates may be linked to population size.
3. Cross-tabulation of organization and activity
4. State Map of respondents by organization type

Overall, for those interested in understanding the findings, it is important to pay specific attention to each state's particular sampling and response issues. Therefore, the information provided in these breakdown by state is crucial to review before beginning any analysis of these data.

**Table 1 -- Breakdown of survey responses by state**

State	Total responses (all levels)	Framework Data production
Alabama	48	18
Alaska	45	33
Arizona	48	34
Arkansas	35	16
California	129	76
Colorado	119	62
Connecticut	45	11
Delaware	36	20
Florida	119	68
Georgia	237	88
Hawaii	31	17
Idaho	82	41
Illinois	189	66
Indiana	89	35
Iowa	97	51
Kansas	155	48
Kentucky	207	47
Louisiana	19	10
Maine	52	22
Maryland	56	27
Massachusetts	190	61
Michigan	197	114
Minnesota	294	155
Mississippi	61	31
Missouri	96	13
Montana	69	46
Nebraska	178	42
Nevada	60	37
New Hampshire	68	35
New Jersey	126	59
New Mexico	70	37
New York	118	31
North Carolina	196	114
North Dakota	16	6
Ohio	152	73
Oklahoma	59	42
Oregon	252	134
Pennsylvania	194	89
Rhode Island	1	1
South Carolina	115	52
South Dakota	14	8
Tennessee	51	21
Texas	254	96
Utah	51	33
Vermont	18	18
Virginia	164	93
Washington	108	84
West Virginia	46	24
Wisconsin	130	107
Wyoming	71	39
<b>TOTAL</b>	<b>5257</b>	<b>2485</b>

**Table 2 -- Breakdown of county-level responses by state**

State	County level responses	County-level responses with Framework Data	Number of counties (From 1990 Census)
Alabama	10	5	67
Alaska	6	6	25
Arizona	14	12	15
Arkansas	3	2	75
California	62	45	58
Colorado	31	26	63
Connecticut	0	0	8
Delaware	7	6	3
Florida	38	26	67
Georgia	39	27	159
Hawaii	9	5	4
Idaho	22	14	44
Illinois	41	28	102
Indiana	25	19	92
Iowa	44	33	99
Kansas	20	16	105
Kentucky	15	9	120
Louisiana	4	2	64
Maine	1	1	16
Maryland	17	12	24
Massachusetts	5	5	12
Michigan	53	38	83
Minnesota	66	47	87
Mississippi	6	5	82
Missouri	19	7	114
Montana	18	15	57
Nebraska	17	11	93
Nevada	12	8	17
New Hampshire	0	0	10
New Jersey	28	18	20
New Mexico	20	14	32
New York	29	20	58
North Carolina	74	62	100
North Dakota	0	0	53
Ohio	65	46	88
Oklahoma	19	15	77
Oregon	40	30	36
Pennsylvania	25	14	67
Rhode Island	0	0	3
South Carolina	26	19	46
South Dakota	2	0	66
Tennessee	9	7	93
Texas	81	48	254
Utah	20	16	29
Vermont	4	4	14
Virginia	48	37	93
Washington	31	29	38
West Virginia	13	7	51
Wisconsin	65	59	72
Wyoming	21	20	23
<b>TOTAL</b>	1224	895	3078

The 26 states chosen for analysis comprise a total of 62% of all responses<sup>1</sup>. Of the 3340 responses, 70% indicated that their organization was involved in creating, integrating, updating, or disseminating digital geographic data. Of the 2323 respondents producing digital geographic data 74% indicate that their organization is involved in creating, integrating, updating, or disseminating framework data.

**Table 3** –Percent Respondents Producing Data by Jurisdiction Type\*

<i>Jurisdiction type</i>	<i>Frequency</i>	<i>Percent of data producing respondents</i>
County	1204	33%
State	789	21%
Municipality	713	19%
Multi-County	498	14%
National	183	5%
Tribal	98	3%
Other	105	3%
Sub-County	51	1%
Multi-State	41	1%
n=3688		

<sup>1</sup> Arizona, Arkansas, Colorado, Florida, Georgia, Iowa, Idaho, Illinois, Indiana, Kansas, Maryland, Minnesota, North Carolina, Nebraska, Nevada, New Jersey, New Mexico, New York, Ohio, Oregon, South Carolina, and Utah.

\* Analysis of 26 selected states only.

## NATIONWIDE COUNTY-LEVEL RESULTS

### County Level Analysis:

Participants in the planning process of the survey placed a particular emphasis on measuring framework activity on the county level because:

- a. it is a commonly established jurisdictional boundary that is already known
- b. it is the smallest such unit that provides a complete coverage of the US
- c. it would be easy for larger regional organizations to use to estimate their jurisdictional boundaries
- d. it was deemed to be a reasonable number (large enough to produce a sizable survey population) but small enough to be affordably surveyed.

This raises a potential problem in that different counties may have different agencies or organizations that are engaged in the generation of framework data. While some counties might have a GIS coordinator who is aware of all data creation activity within their county, others might have a surveyor creating geo-reference data that is unaware of a conservation agent creating hydrography data. If the state coordinator only contacted one of these data producers, the survey would probably be an incomplete reflection of the activity in the jurisdiction.

To demonstrate the importance of recognizing the limitations in analyzing all 50 states, we conducted an analysis of county level responses for those states with high response rates by county. In other words, we examined each state's response rate for based on a) total number of counties within a state, b) number of counties surveyed, c) number of county level responses and finally d) percent of responded counties who engage in framework activity (Table 1). As evident, first the numbers of counties within a state vary considerably based on state size and history. Second the percent of counties surveyed within a state varied as well. And finally the response rate of counties by state also varied considerably. Taking all these factors together, 24 states were eliminated from the subsequent analysis of county organizations. These 24 states were eliminated based on 2 main criteria: 1) within state response rate of less than 50% of all organizations surveyed; 2) states with biased response rates, for example response rates associated with population county size. Because it is extremely difficult to manipulate these data for a worthwhile analysis, we felt that focusing on 26 states was more practical.

Based on this sample of 26 states, respondents who indicated that they were NOT in any way involved with Framework data were eliminated from this analysis. As a result, 836 county level responses (out of a nationwide group of 1600 county responses) are analyzed. It is important to recognize that the 836 responses analyzed here within the 26 high-response states is not generalizable to national level comparisons (statements) of Framework activity. These data represent only the 26 states included in this sample and due to the numerous limitations in the overall nationwide sampling design, it is not valid to draw larger conclusions. In other words, the responses discussed here, while representative of the 26 states in the sample, are not validly representative of all 50 states. Readers must be careful not to over-generalize the results discussed here and recognize the methodological limitations of the study overall. Nevertheless, the findings from the states included here significant and provide to begin a portrait of framework activity throughout the different regions of the country. Except where noted, the following analysis represents 836 county level responses.

The department(s) at the respondents' agency currently engaged in GIS work from a list of 20 possible departments was measured: "Within your organization, please indicate which

departments (or groups) are creating, maintaining, inventoring, distributing, or using digital geographic data." The possible departments include the following:

- GIS/Mapping
- Land Records (Assessor, Recorder, Clerk, etc.),  
Community Development (Planning, Zoning,  
Economic Development, etc.)
- Public Safety (Fire, Police, etc.)
- Transportation
- Water
- Wastewater
- Other Public Works (Excluding Water, Wastewater,  
Transportation)
- Engineering (Surveying, Street Lights, etc.)
- Health and Human Services
- Elections, Education (School Districts)
- Libraries
- Administrative & Finance
- Information Services (Data Processing)
- Environmental
- Historical Preservation & Archeology
- Natural Resources
- Agriculture

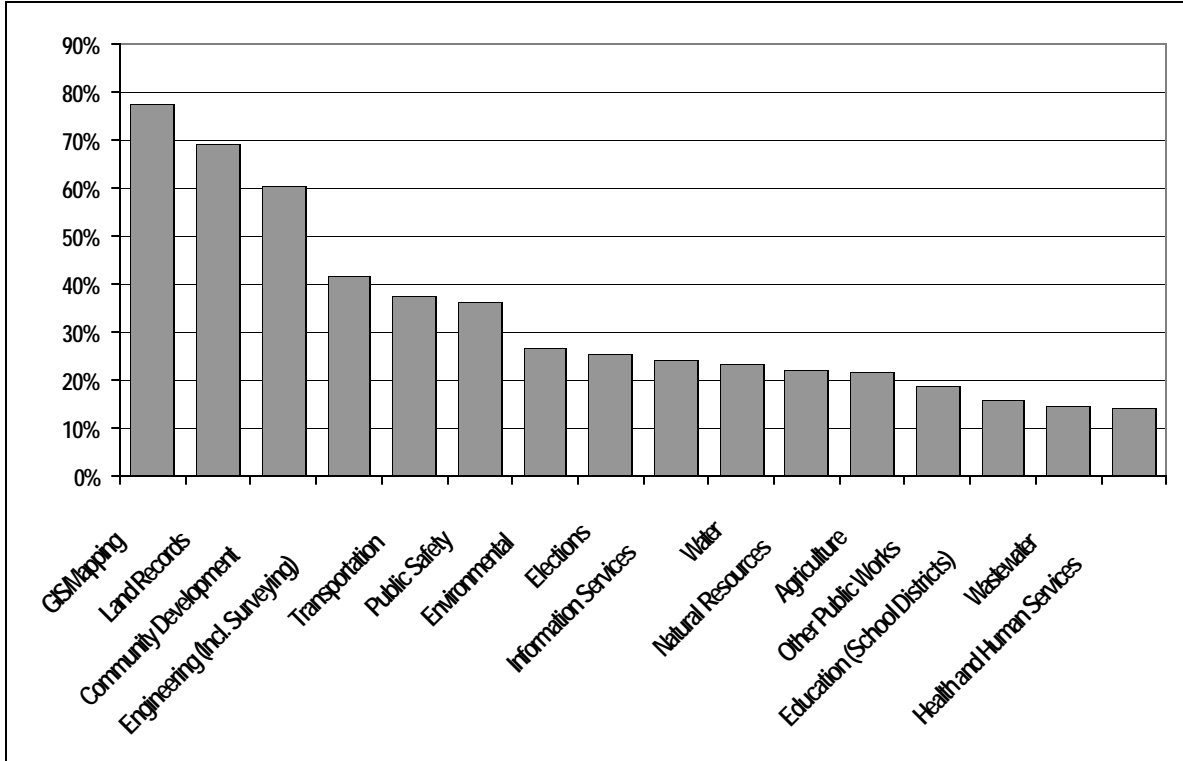
The vast majority of respondents are using multiple agencies for their counties' GIS work (Table 4). More than half of the respondents (61%) responded that 4 or more departments within the county agency are engaged in framework GIS work. This reflects an important change in GIS organizations. The pattern, of multiple departments engaged in GIS work within a single agency, suggests that the traditional Multipurpose Land Information System concept (Brown and Moyer 1989) is supported by these data.

Another significant finding from these results is that 77% of respondents are using a GIS or mapping office to conduct their GIS work. The GIS/Mapping department was the single most common response regarding the question of department activity in GIS work. This finding suggests that GIS work has become so important for many county agencies that GIS/Mapping departments are created to serve this need. The implication is therefore; GIS has arrived in local government.

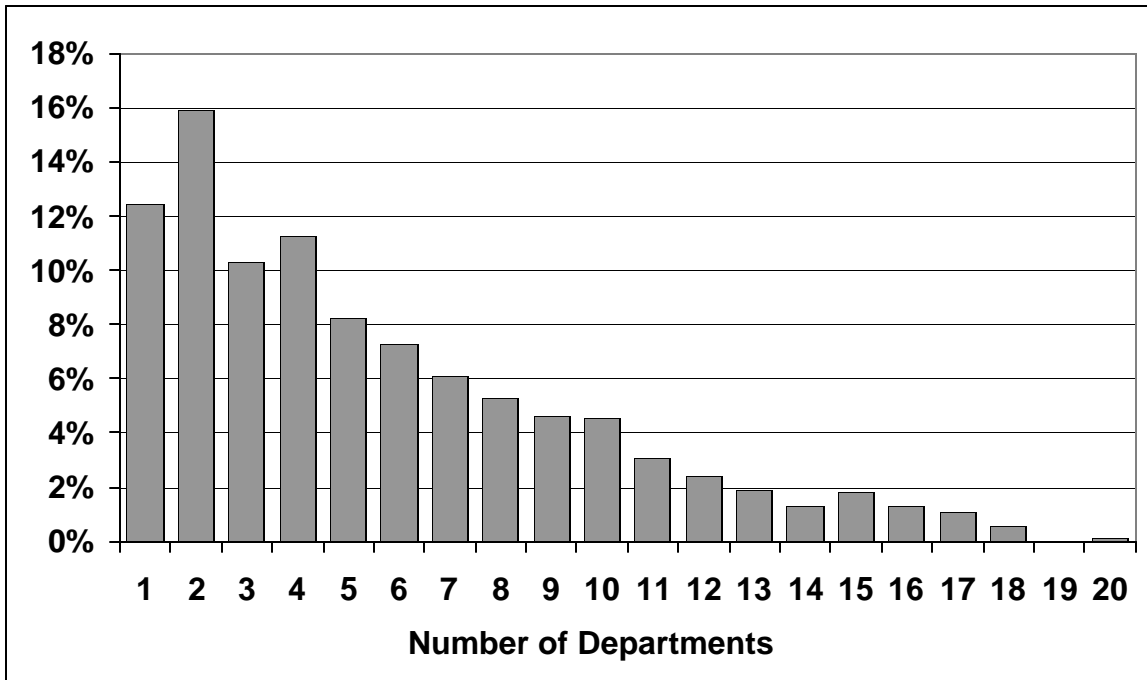
**Table 4** -- Based on responses to Question 13, this shows a count of the number of respondents that have GIS work occurring in varying numbers of agencies.

<b>Number of Agencies Using GIS</b>	<b>Number of Responses</b>	<b>Percent of Total Number of Responses</b>
1	104	12%
2	133	16%
3	86	10%
4	94	11%
5	69	8%
6	61	7%
7	51	6%
8	44	5%
9	39	5%
10	38	5%
11	26	3%
12	20	2%
13	16	2%
14	11	1%
15	15	2%
16	11	1%
17	9	1%
18	5	1%
19	0	0%
20	1	0%

In addition, the vast majority of respondents are using multiple departments or agencies for their counties' GIS work (Figure 3). 61% of respondents reported that four or more departments within the county agency are engaged in Framework GIS work. This reflects an important shift in GIS organizations from single agencies to multi-agencies, supporting the traditional Multipurpose Land Information System (MPLIS) concept (Brown and Moyer 1989). The concept of the MPLIS focuses on systems consisting of multiple departments or groups constructed with shared (often transactional) data used for a variety of purposes. In the future, therefore, it is likely that GIS work will continue to be shared among multiple departments. What this does not tell us, however, is the mechanisms by which data is utilized and shared among departments within a single organization.



**Figure 2 --** Percent of respondents employing specific organizations for GIS work



**Figure 3 --** A graph of the number of departments actively using GIS

An additional measure of resources comes from the following question about staff size: "Please estimate the range that best reflects the number of full time employees (equivalents) in your organization currently involved in development, inventory, coordination, integration, or distribution of framework data." 62% indicated having between one and five full time employees (Table 5). Almost 20% of respondents indicated that they had the equivalent of less than one full time employee. This could mean that the only staff member working on framework has time split between these GIS-related activities, and some other non-framework tasks (which might include GIS work on non-framework themed data). Overall, these data demonstrate that digital data users may be present within multiple departments within the county organization as well as indicate that more than one employee may be engaged in framework data work.

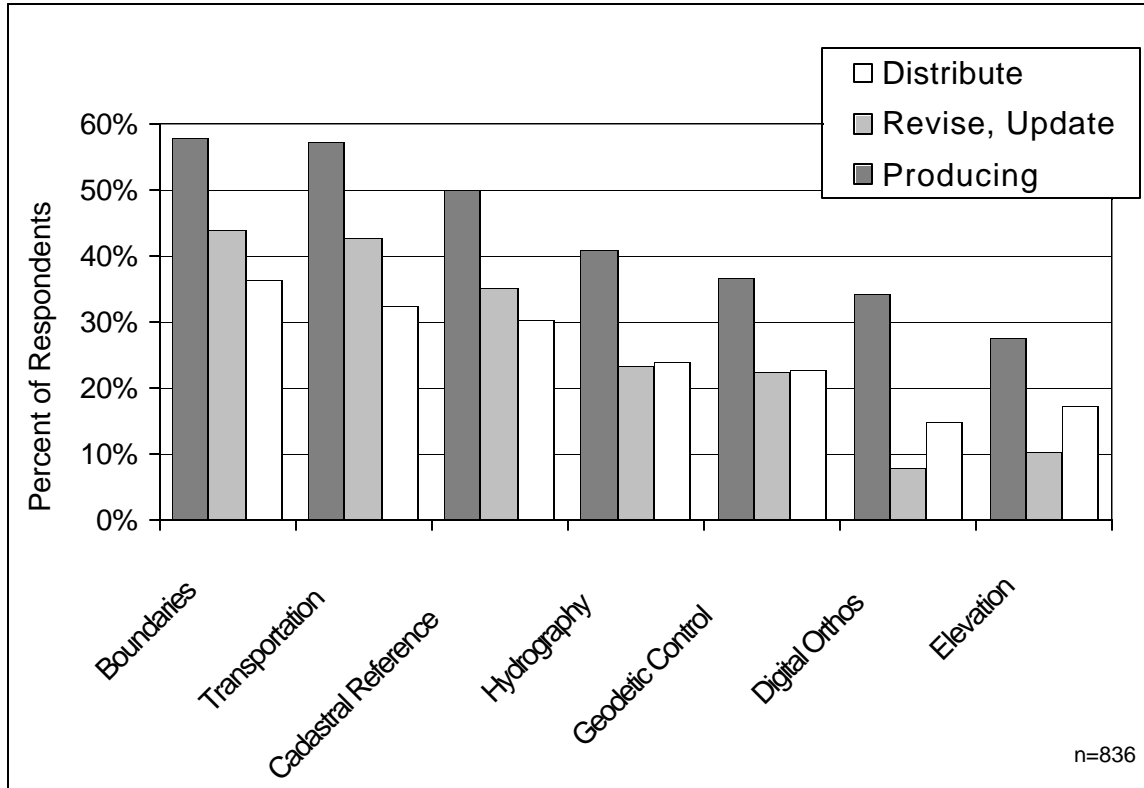
**Table 5 – Percent Data Producing Organizations by Staff Size\***

<i># Full Time Staff</i>	<i>Frequency</i>	<i>Percentage</i>
< 1	119	18%
1 - 5	415	62%
6 - 10	82	12%
11 - 20	39	6%
> 20	19	3%
n=674		

\* A response of less than one suggests staff members contributing on a part time basis.

Respondents were asked whether they produced each of the 7 separate Framework themes. For each theme an additional question was then asked: "Do you create, update, integrate or distribute" framework data specific to each of the individual themes. The theme of governmental units was the most highly reported theme activity with 58% of respondents stating that they collect geospatial data on government boundaries (Figure 4). As expected, a similar finding was reported for transportation data, with 57% of respondents responding that they collect transportation data. Only 28% of respondents reported collecting elevation data, which may reflect a heavy reliance on federally-produced elevation data. In addition, 34% reported digital orthoimagery, perhaps because it can be very costly and is only rarely done in small sets.

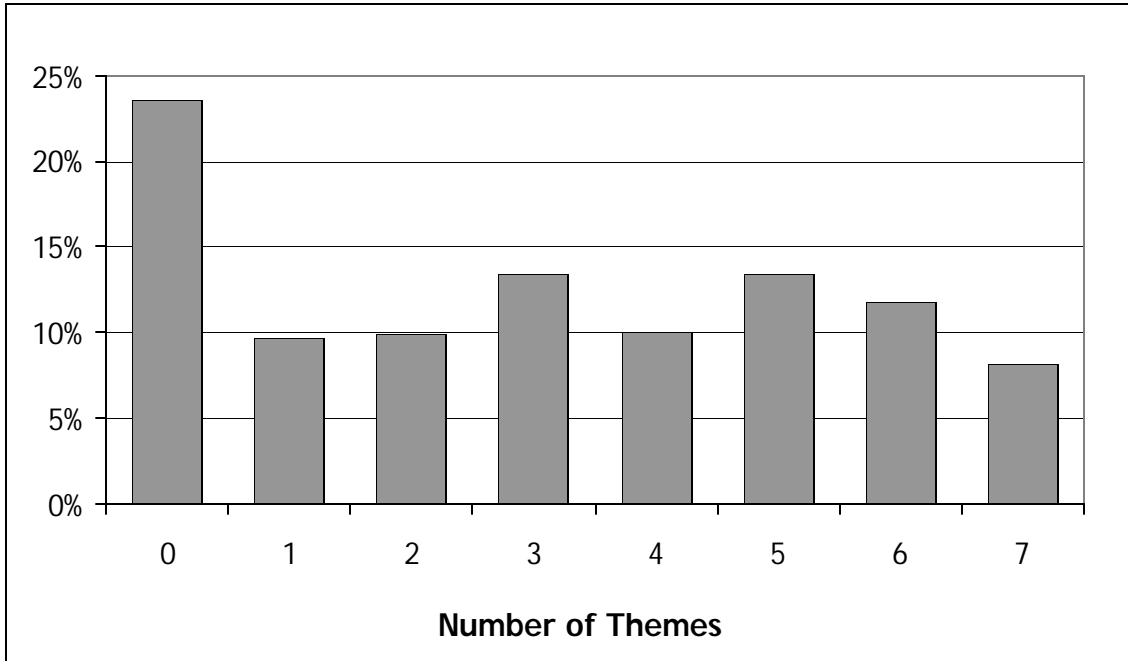
A separate set of questions inquired about other activities revolving around the individual framework data themes. For the more common data themes -- particularly government boundaries, transportation, cadastral reference -- a significant portion of the respondents actively producing thematic data were also revising and updating those data themes. This offers promise for systems in which maintenance has become a planned part of the data production process. Other data themes -- particularly elevation and digital orthoimagery -- which are seldom produced on an in-house basis had much lower levels of being revised or updated, but these themes were distributed to others (Figure 4).



**Figure 4** — Percent of respondents engaged in 3 different activities with each framework theme. Some data themes, particularly elevation and digital orthoimagery, are distributed more commonly than they are updated.

Figure 5 illustrates the percent of respondent organizations creating zero to seven themes. While a considerable portion, 24%, of respondents indicated that they were not producing any framework data themes, over 30% indicated that their organization was producing five or more different framework themes. There is considerable variability in county organizations production of specific themes, ranging from zero themes to 8% of organizations creating all seven themes.

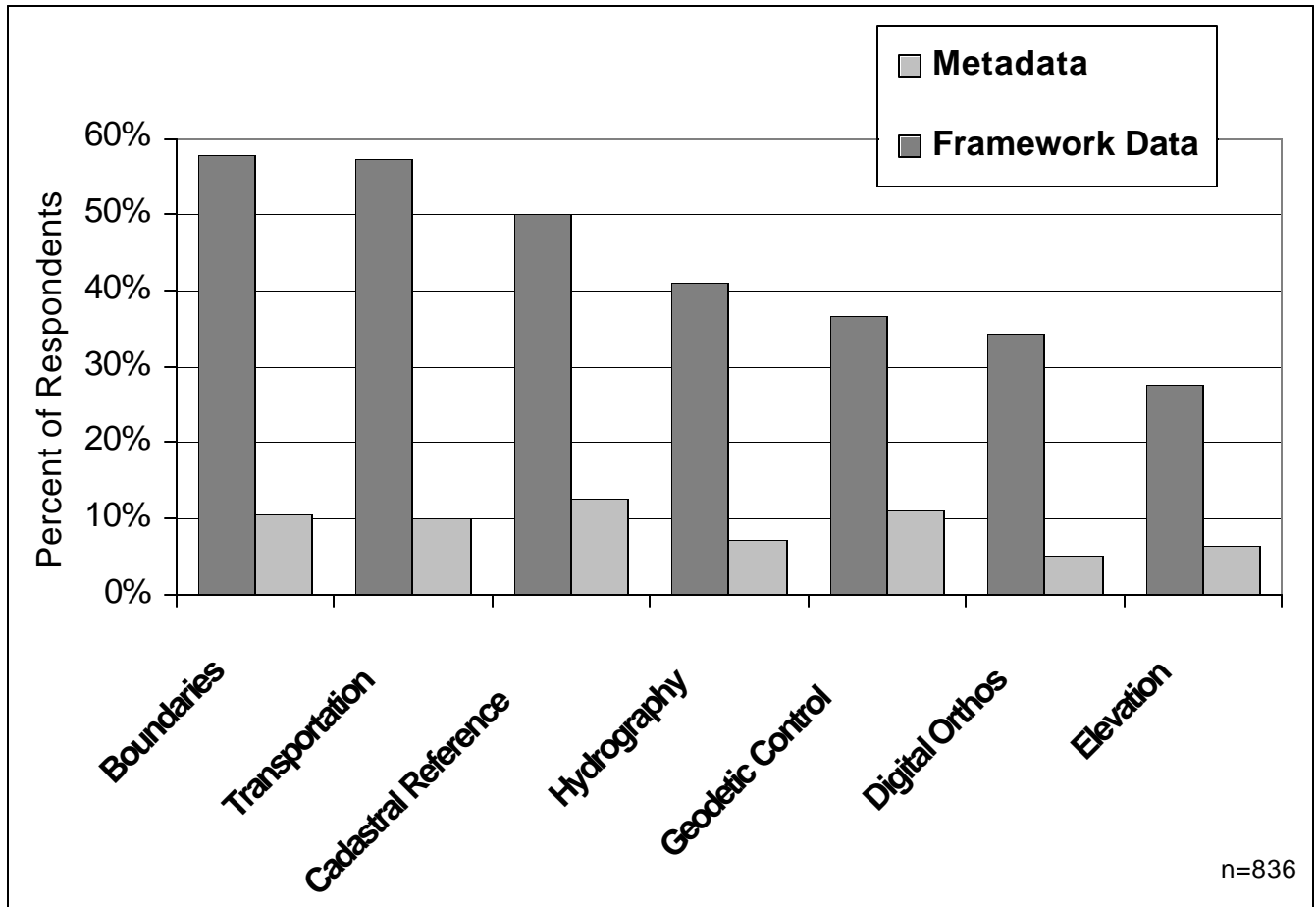
Respondents who indicated activity in any of the seven framework themes were also asked whether they maintained metadata for that theme. Metadata is specific information explaining how data has been produced and maintained. It also includes contact information for the person(s) responsible for the data. Overall, a large number of respondents did not maintain metadata for any of the framework themes in which they produce data (Figure 6). Specific themes were found to be more likely than others to have metadata maintained. For example, only 15% of the respondents who were producing digital orthoimagery were maintaining metadata. In contrast, 30% of the respondents producing geodetic control were maintaining metadata. These findings raise questions about whether certain themes have low metadata development rates because of the difficulty of developing the metadata or because of a lack of need in using the metadata.



**Figure 5** – Number of framework themes created, updated, integrated, or distributed by respondent organizations.

### Data Dissemination

Much of the NSDI’s work is based on the exchange or sharing of data. To explore the readiness of county level jurisdictions for sharing data, the survey measured several issues regarding different aspects of exchanging or sharing data. Respondents were asked, "Does your organization share your data with other organizations?" 88% of county level organizations stated that they did share data (Table 6). This finding serves as a relatively basic measure of an organization’s ability and willingness to allow other data users to access their data. However, the wording of this question does not measure the mechanisms behind sharing practices and whether these sharing activities occur with other caveats or conditions attached. Thus, among the population of organizations that are sharing data, there might be restrictions including, charging for data access or allowing some agencies access to data while denying others access. A more specific measure of sharing data needs to be investigated in order to fully understand agencies’ sharing policies and practices and the impact on creating the NSDI.



**Figure 6** – Percent of organizations producing framework data and metadata whose organizations are active in the production of 7 different themes of framework data (dark gray) and those who were actively maintaining metadata for each of the themes (light gray).

**Table 6** – Percent Organizations Sharing Data

<i>Data Sharing</i>	<i>Frequency</i>	<i>Percent</i>
Share Data	707	88%
Do Not Share Data	99	12%
n=836, 30 missing responses		100%

A specific example of the complexities of data sharing is reflected in the various ways in which organizations institutionalize data sharing policies. The survey included a measure of redistribution of data, "Does your organization permit others to redistribute its data?" Only 30%

of organizations sharing data allowed the unrestricted redistribution of their data (Table 7), while the majority, 70% of organizations implemented some form of restricted redistribution. 25% indicated that they did not allow any redistribution. Another 40% indicated that they allowed redistribution under certain restrictions. While some of these restrictions might simply include a requirement about crediting the original source of the data, others could be strict to the point of preventing a free flow of data. Since this number includes some non-governmental and private organizations, this is not an unreasonable expectation. Redistribution is going to continue to be a significant issue because data and information are commodities that can be easily reproduced. As participation in NSDI increases, sharing and redistribution issues will become more important. A limitation of the survey data, however, is that there are no details about sharing requirements.

**Table 7 – Percent Organizations Redistributing Data**

<i>Redistribution Practices</i>	<i>Frequency</i>	<i>Percent</i>
Yes, the data can be redistributed by others, with no restrictions by us	232	30%
Yes, the data can be redistributed by others, but only under conditions set by us	350	45%
No, the data may not be distributed under any condition	196	25%
n=836, 58 missing responses		

A common mechanism for encouraging sharing and more generally data development, is the coordinating council. Respondents were asked, "Does your organization participate with a geographic data coordinating council or group?" 42% of the respondents indicated participation in such a coordinating group (Table 8). 58% of the respondents indicated that their organization did not "participate with a geographic data coordinating council or group." Considering the relatively little cost or effort required to participate in one of these groups, the low response rate raises serious concerns about the underlying causes for the low participation rate.

**Table 8 – Percent Organizations in Coordinating Council**

<i>Coordinating Council</i>	<i>Frequency</i>	<i>Percent</i>
Yes	339	42%
No	470	58%
N=836, 27 missing responses		

When asked "to list the two geographic data coordinating councils or groups" with which respondents were most active, a fairly robust list was produced. The open-ended question made it difficult to analyze in detail, but a descriptive analysis indicates that 50% of respondents listed a state-level organization. Examples include the Wisconsin Land Information Board, the New Jersey State Mapping Advisory Committee, and the Colorado Counties GIS/LIS Committee. While a few respondents mentioned national organizations (including the census bureau and URISA), the remainder were mostly regional and local organizations. Some of these coordinating councils included formal institutions with some authority over participants, like MetroGIS or the Cincinnati Area GIS (CAGIS). Others were voluntary organizations (like the Eastern Oregon GIS User Group) which were more loosely bound together professionals working in a variety of settings, including the public and private sector. The predominance of state organizations indicates both the success of such groups and the regional perspective shared by many local data producers.

A more systematic method of data sharing can be insured with a specific policy, however this was not found to be as common as might be expected. Only 40% of the survey respondents indicated that they had a policy describing data distribution, while 60% said that they had no policy (Table 9). This suggests that many of the organizations that are actively disseminating data are doing so without any formal policy describing whom to share data with, the conditions of the data sharing, and the liability for any problems with the data.

**Table 9** – Percent Respondents with Policy on Data Dissemination

<i>Dissemination Policy</i>	<i>Frequency</i>	<i>Percent</i>
Yes	320	40%
No	488	60%
N=836, 28 missing responses		

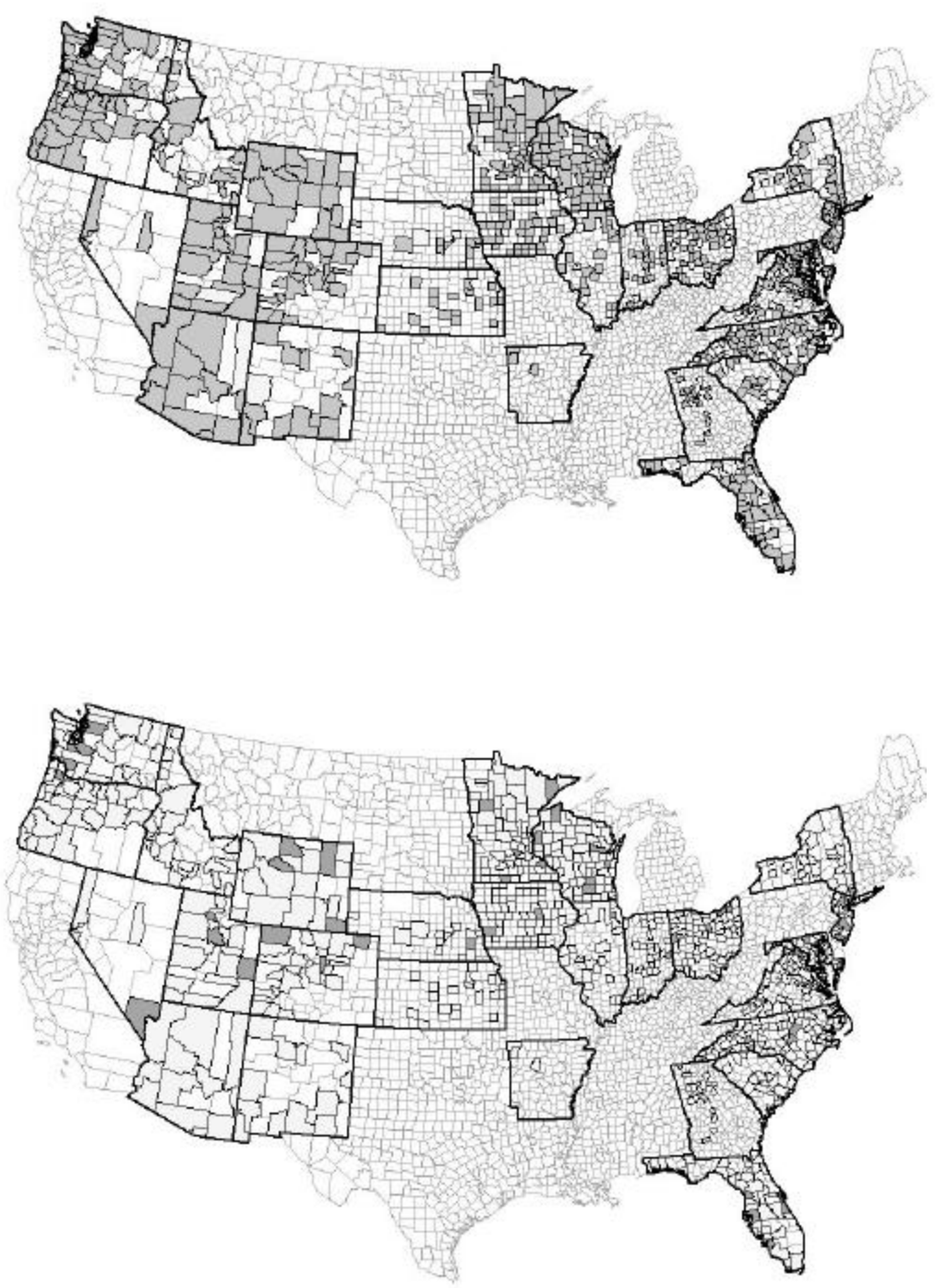
Analysis shows that the potential role of data sharing policy is related to whether or not organizations do engage in sharing data (Table 10). Responses to the survey indicate that having a data sharing policy (even though some policies prohibit sharing) increases the likelihood of an organization sharing data with other organizations. Specifically, 95% of respondents whose organizations had data sharing policies, indicated that their organization was sharing data. In contrast to those without a policy, only 78% were actively sharing data with other organizations. While this may not seem surprising, it shows how significant the formalized nature of policies can be to ensure specific organizational activities.

**Table 10** – Percent Organizations Sharing Data by Data Sharing Policy

	<b>Data Sharing Policy</b>		
<b>Sharing Data with Other Orgs</b>	Yes	No (or N/A)	
Yes	304 (95%)	403 (78%)	
No (or N/A)	16 (5%)	113 (22%)	
Total	320 (100%)	516 (100%)	
Chi-Square = 43.225, df = 1, p<.05			

**Table 11** – Percent Respondents Advertising Data

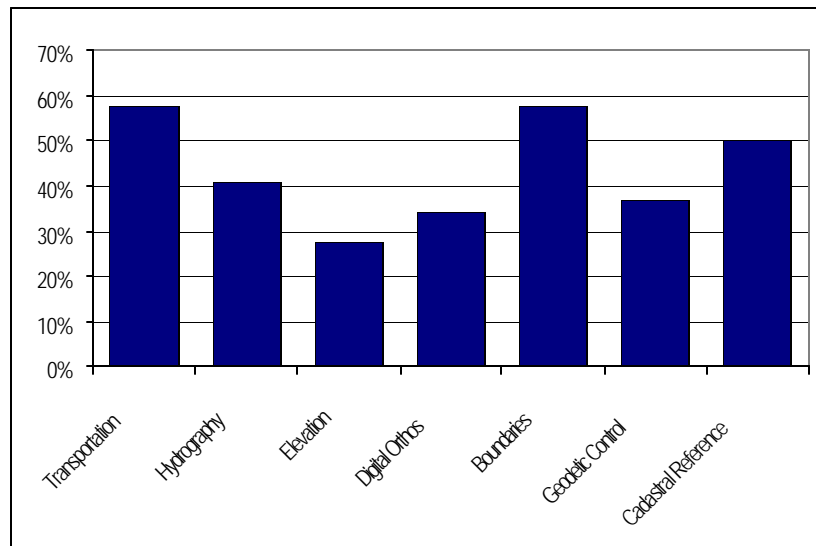
<i>Advertise Data</i>	<i>Frequency</i>	<i>Percent</i>
Yes	69	9%
No	728	91%
N=386, 39 missing responses		



**Figure 7** -- While many counties share data with other organizations (707 shown in dark gray in the map on top), few advertise their data in catalogues or clearinghouses (66 shown in dark gray in the bottom map). In both maps, all other responses (No or not applicable) are shown in light gray.

Another interesting result regards the sharing of public data. In order for the FGDC's Framework concept to be realized many agencies will need to share their geospatial data with other organizations. Fortunately for the FGDC, 85% (707) of county respondents indicated that they do share data (Figure 7). However, the ways in which they share may indicate some issues that remain to be addressed. Less than half (38% or 320) have a policy describing how they should share data. And still fewer counties (66 or 7%) actively advertise their data on catalogues or clearinghouses. This final finding should be considered important to the FGDC in regards to their Clearinghouse effort.

Respondents were asked whether they were involved with each of the 7 Framework themes separately. Each question asked, "Do you create, update, integrate or distribute framework transportation data?" The theme of governmental units was the most highly reported theme activity with 58% of respondents reporting that they collect geospatial data on government boundaries. A similar finding was reported for transportation data, with 57% of respondents reporting that they collect transportation data. The constant changing nature of transportation networks, makes locally-produced data valuable. Only 28% of respondents reported collecting elevation data which may reflect a heavy reliance on federally produced elevation data.



**Figure 8** — The percent of respondents involved with each framework theme.

In addition to collecting data on the 7 framework themes, there are measures by theme of accuracy, cost, staffing, specific thematic elements, and metadata. For example, the accuracy of county-level elevation data might provide further insight into the small percentage of county agencies collecting elevation data. Of the 231 county level respondents who reported that they are involved elevation data, 45% of them described the absolute vertical quality of their data as better than 4 foot contours (or its equivalent). In addition, 68% indicated that they are currently distributing their elevation data. However, only 26% of the county respondents maintain metadata for their elevation data and only 5% report they are FGDC compliant.

## TRANSPORTATION FRAMEWORK DATA FINDINGS

These data have the potential to be analyzed based on responses about each of the seven framework themes. As an example of such analysis, this section offers an analysis of responses about transportation framework issues. However, this report does not include similar analysis of other framework themes.

### Transportation Framework Data Responses

Out of the selected sample of 24 states, 1220 organizations indicate specifically that they are producing transportation framework data. Of the 1220 organizations producing transportation data, almost 40% are organizations whose jurisdiction is most accurately described as a county (Table T1). 21% of transportation data producers are municipalities, 14% are multi-county organizations and 13% comprise state-level organizations. County level organizations are the largest group of transportation data producers based on the findings in this survey.

**Table T1**– A frequency table displaying the types of jurisdictions that are represented among the survey responses.

<i>Jurisdiction Type</i>	<i>Frequency</i>	<i>Percentage</i>
COUNTY	480	39%
MUNI	254	21%
MULTI-COUNTY	172	14%
STATE	161	13%
NATIONAL	55	5%
OTHER	35	3%
TRIBAL	38	3%
MULTI-STATE	15	1%
SUB-COUNTY	7	1%

N=1217, 3 missing responses

The survey measures separately the type of transportation data that organizations keep: road, railroads, waterways, trails, bridges, airports, ports, and tunnels. Nearly all (99%) of the 1220 transportation data producers are involved in the production of “roads” data (Table T2). This finding was expected since many organizations consider “roads” as the “base layer” of either their transportation data or their entire database. Next, 67% of organizations produce railroads data and 62% produce waterways data. 50% of organizations report that they kept trails data as part of their transportation data, a surprising finding especially considering that this data is relatively difficult to collect, carries less fiscal significance, and are often somewhat transitory. 44% of organizations keep bridge transportation data and 43% keep airport data. Finally, ports, 10%, and tunnels, 9%, are reported infrequently, perhaps reflecting both the less common occurrence of these in the built landscape and a low level of interest in these features.

**Table T2** – Table of the responses to the question, “The transportation theme includes sub-categories. Please indicate what type of data is included in your transportation framework data.” The question allowed the selection of as many options as necessary. 5 respondents made no selections.

<i>Progress of Work</i>	<i>Percent Yes</i>	<i>Percent No</i>	<i>Total</i>
Roads - Transportation Framework	99%	1%	100%
Railroads – Transportation Framework	67%	33%	100%
Waterways - Transportation Framework	62%	38%	100%
Trails - Transportation Framework	50%	50%	100%
Bridges - Transportation Framework	44%	56%	100%
Airports - Transportation Framework	43%	57%	100%
Ports - Transportation Framework	10%	90%	100%
Tunnels - Transportation Framework	9%	91%	100%

N=1220

In addition to questions about the thematic contents of their data, respondents were asked, “If your organization creates or revises / updates road data, does it include street addresses or linear reference system information (e.g. mile point, reference post, or engineering stationing)?” Of those organizations keeping road data, 54% indicate that their data include street addresses while 40% indicate that they include linear reference information (Table T3). While it might be assumed that organizations with linear reference information would also keep street addressing data, the survey found otherwise. Of the 891 respondents keeping either type of data, 74% keep only one type of data while only 26% keep both types of data. This relationship may reflect the unique variety of organizational duties associated with each type of data.

**Table T3** – Table of the additional reference data associated with road data.

<i>Additional Road Data</i>	<i>Frequency</i>	<i>Percentage</i>
“Street address” only	412	34%
“Linear reference system information” only	247	20%
Both “Street address” and “Linear reference system information”	232	19%
No Answer	329	27%

N=1220

Of those 1220 respondent organizations who “create, update, integrate or distribute” transportation data, 71% indicate that they “currently” create data (Table T4). 11% of organizations respond that they plan to create transportation data in the future. And 18% report no current activity and no plans to create transportation data and therefore are updating, integrating or distributing transportation data only.

**Table T4** -- Table of the responses to the question, “Does your organization create transportation data?”

<i>Progress of Work</i>	<i>Frequency</i>	<i>Percentage</i>
Currently create	863	71%
Plans to create	132	11%
No plans to create	217	18%

N=1212, 8 missing responses

Next, of all the organizations currently creating transportation data, only 31% indicate that their transportation data is complete (Table T5). In contrast, a majority of organizations, 65%, indicate that the transportation datasets they create are considered works in progress. These findings suggest one of the common limitations for transportation data users relying on other.

**Table T5** -- Table of the responses to the question, “If your organization creates transportation data, to what level has your organization created digital transportation data covering your service or jurisdiction area?”

<i>Progress of Work</i>	<i>Frequency</i>	<i>Percentage</i>
Completed	268	31%
Work in Progress	558	65%
Planned	35	4%

N=861, 2 missing responses

57% of organizations engage in the distribution of transportation data, 22% of organizations plan to distribute transportation in the future, and 21% have no plans to distributed transportation data (Table T6).

When asked, “Does your organization coordinate transportation data creation with other organizations?” 44% indicate they currently do create data with other orgnaizations, 26% report plans to do so and 30% indicate having no plans to do so (Table T7).

Finally, 40% of orgnizations indicate that they are currently engaged in integrating transportation data with other organizations’ transportation data, 32% plan to integrate in the future and 27% have no such plans to integrate transportation data (Table T8).

**Table T6** – Table of the responses to the question, “Does your organization distribute transportation data to other users?”

<i>Distribute Data</i>	<i>Frequency</i>	<i>Percentage</i>
Currently distribute	688	57%
Plans to distribute	264	22%
No plans to distribute	258	21%

N=1210, 10 missing responses

**Table T7** – Table of the responses to the question, “Does your organization coordinate transportation data creation with other organizations?”

<i>Coordinate Data</i>	<i>Frequency</i>	<i>Percentage</i>
Currently coordinate	535	44%
Plans to coordinate	311	26%
No plans to coordinate	365	30%

N=1211, 9 missing responses

**Table T8** – Table of the responses to the question, “Does your organization integrate transportation data with other organizations' transportation data?”

<i>Integrate Data</i>	<i>Frequency</i>	<i>Percentage</i>
Currently integrate	488	40%
Plans to integrate	392	32%
No plans to integrate	331	27%

N=1211, 9 missing responses

### **Transportation Data Quality**

While the data described above indicate that many organizations currently deal with transportation data, it is also very important to measure the quality of transportation data. Data quality was measured by the following question, “Which of the following best represents the approximate positional accuracy of the majority of your transportation data?” 39% of organizations indicate that their organization’s data are of the highest category of accuracy, more accurate than 1:2,4000 scale (Table T9). Next, 11% of organizations reported the next highest level of accuracy reflecting source material whose scale ranged from 1:2400 to 1:5999 and 21% indicate a scale range that would include 1:24000 USGS topographic quad maps as source material. Only 9% of organizations indicate a less accurate positional accuracy representing 1:25000 or lower. Finally, 19% of organizations who keep transportation data indicate that they are unsure of the positional accuracy of their transportation data. This finding may simply reflect a lack of understanding of general accuracy issues or it may reflect a specific lack of knowledge due to data developed under a clouded history. Further investigation is needed to determine the reasons why 19 % of organizations are ‘not sure’ of the accuracy of their transportation data.

**Table T9** – Table of the responses to the question, “Which of the following best represents the approximate positional accuracy of the majority of your transportation data?”

<i>Positional Accuracy of Transportation Data</i>	<i>Frequency</i>	<i>Percentage</i>
More accurate than 1:2,400 scale	472	39%
1:2,400 to 1:5,999 scale	138	11%
1:6,000 to 1:25,000 scale	256	21%
Less accurate than 1:25,000 scale	111	9%
Not sure	230	19%

N=1217, 3 missing responses

### **Transportation Metadata**

Another important measure of the quality of transportation data is the maintenance of metadata. Of respondents producing transportation data, only 21% are currently maintaining metadata (Table T10). Many more respondents, 37%, are planning to maintain metadata, and 40% have no plans to maintain transportation metadata. Despite the recent visibility of metadata efforts, it is not surprising that metadata maintenance is still at a relatively low rate considering both the lag time in the survey analysis and the time required to institutionalize metadata development processes. In addition, this low level of metadata may be a factor in the relatively high number of respondents who do not know the accuracy of their transportation data.

Since much of the promotion of metadata in the United States has been done by the FGDC, one might anticipate that a great deal of the metadata would be compliant with the FGDC content standard for metadata. However, only 25% of those organizations maintaining metadata indicate that their metadata is FGDC compliant (Table T11). In contrast, 46% indicate that they are unsure whether their metadata was compliant. This may simply reflect that some respondents are unfamiliar with the content standard however, it also seems likely that the complexity of the content standard is perhaps a significant factor underlying this high percentage of organizations who do not know if their metadata is compliant.

Finally, a large majority of organizations, 77% of those maintaining transportation metadata are also engaged in the dissemination of transportation data. This is a positive finding since one of underlying purposes of metadata is dissemination.

**Table T10** – Table of the responses to the question, “Does your organization create and maintain transportation metadata?”

<i>Metadata</i>	<i>Frequency</i>	<i>Percentage</i>
Currently create	257	21%
Plans to create	452	37%
No plans to create	490	40%

N=1211, 9 missing responses

**Table T11** – Table of the responses to the question, “If your organization has created transportation metadata, is it compliant with the FGDC Content Standard for Digital Geospatial Metadata?”

<i>FGDC Compliant</i>	<i>Frequency</i>	<i>Percentage</i>
Yes	62	25%
No	74	29%
Not Sure	115	46%

N=251, 6 missing responses

### **Transportation Data Investment**

As a measure of transportation data investment, respondents were asked about annual expenditures on transportation framework data. Although this does not provide a very precise measure of data quality, annual investments may suggest variations in institutional commitments to transportation data. The measure is further limited by use of a five-point scale, with the highest category only measuring annual expenditures of \$500,000 or greater towards “creating, updating, integrating, and/or distributing.” A very small percent, 3%, of organizations indicate that they spend an estimate of \$500,000 or higher each year. This category represents a minimal combined estimate of \$17.5 million annual investment (Table T12).

However, it is easy to imagine that one of the responding federal agencies could spend millions more than that in a given year. Assuming the minimum amounts for each response category, the survey represents \$32 million in transportation data production investment. For each range of investments the only amount of which we can be certain is the minimum, so the minimum is taken as a particularly conservative measure of annual investment. Readers should note that more than half of the respondents indicated an investment of “< \$10,000” which is included in the minimum combined investment as zero expenditures. An expected finding is that the majority 60% of those organizations spending the largest amount of money are states jurisdictions and 18% of organizations are national jurisdictions (Table T13).

**Table T12** – Table of responses to the question, “Please estimate the amount your organization spends annually on creating, updating, integrating, and/or distributing transportation data?”

Annual expenditures on Transportation Framework			Minimum Combined Investment
	Count	Percent	
< \$10,000	630	54%	
\$10,000 - \$49,999	333	29%	\$3,330,000
\$50,000 - \$99,999	94	8%	\$4,700,000
\$100,000 - \$500,000	72	6%	\$7,200,000
> \$500,000	35	3%	\$17,500,000

N=1164, 56 missing responses

**Table T13** – Of the 34 respondents indicating that their organization invested over \$500,000 in transportation framework annually, the majority were state jurisdictions.

<i>Jurisdiction Type</i>	<i>Frequency</i>	<i>Percentage</i>
STATE	16	60%
NATIONAL	9	18%
MULTI-COUNTY	2	8%
MUNI	3	5%
COUNTY	3	5%
OTHER	1	2%
TRIBAL	1	2%

N=34

Almost every one of these transportation data producers includes roads in their data, but the majority also include railroads, waterways and trails in their data. And the majority are distributing their data to others. These indicators suggest that transportation data is an increasingly available form of framework data.

Many producers have incorporated street address information as part of their roads data. Many have incorporated linear reference system information in their roads data. But, very few have incorporated both. This may indicate an interesting dichotomy of interests between data producers, those interested in events occurring on a roadway (or other linear transportation systems) and those interested in things occurring at buildings near roadways. It also worth pointing out that there is no measure of the quality of these data and, therefore, no means for determining their appropriateness for larger data applications (e.g., Fletcher et al. 1998).

Still some findings suggest barriers to sharing and the compilation of these data as a single, comprehensive data source. Most data producers reported their organization has incomplete data. Many reported that they do not coordinate the development of their data with other organizations and do not integrate their data with other data sets. Most data producers are not developing transportation metadata, and most that are producing metadata are not producing FGDC compliant metadata. And the majority of data producers are investing fairly limited amounts in their data.

## FUTURE APPLICATIONS OF THE FRAMEWORK SURVEY

### Creation of Data Producer/User Phone Book:

One of the intended uses of the survey results is as a directory of likely geo-spatial data producers and users. This may, in fact, be the single most appropriate use of this database. This "GIS phonebook" application of the survey responses holds value for the widest variety of members of the GIS community: individual users can use it to search for datasets, state coordinators can use it to develop a list of likely members of a state-wide effort, and commercial vendors can use it as a marketing tool.

Because the data is available in multiple formats, the "phonebook" can be accessed in multiple ways. Perhaps the simplest is simply to use the ArcView Project (from [www.fgdc.gov](http://www.fgdc.gov)), select the data theme that interests you most and zoom in and the region in which you are interested. Then potential data sources can be identified with simple queries. If a data user were interested in a small metropolitan area, they would need to be careful about considering larger data producing agencies who are geographically located in a place quite far from the study area. Imagine a project manager in Reno, NV who would not only look for local government data producers within a small radius of Reno (say, 20 miles) but would also look throughout the states of Nevada and California for regional and state-level producers, and might even look at the entire national set of federal data producers.

Because it functions as a database program, ArcView also allows for some Boolean queries like, "What are the counties in California with elevation data of an accuracy greater than 1:24000?" Another manner in which the data can be accessed is to open the .dbf file in whatever database program or statistical package that you use. Often, these powerful programs would allow fairly sophisticated queries or selection of producers.

In each of these cases, the database includes contact for information for the survey respondents. While sometimes incomplete, phone number, mailing address, and email are often available. This permits data users to contact potential data producers to explore whether appropriate data are publicly available for use. For state coordinators, this database is already paying dividends (see section of the follow-up survey). The ability to develop a preliminary list of producers and users in their state is allowing some states to jump into the coordination activities. For the FGDC this list has the potential to serve as a starting list of NSDI participants.

Despite the fact that this database is still not a complete list of data producers (something that is essentially impossible) it is an enormous resource for those willing to take the time to seek out potential contacts.

### Follow-up Survey:

In order to examine the possibility of repeating the Survey, a Framework Data Survey Questionnaire was designed by the FGDC/NSGIC and distributed to all fifty state coordinators. The survey questions and results are reported at:

[http://www.fgdc.gov/framework/survey\\_results/update.html](http://www.fgdc.gov/framework/survey_results/update.html). A summary of the results is included in the *URISA Proceedings* paper in Appendix B.

Of the fifty states, forty-six state coordinators responded to the survey. 70% of the state coordinators reported that they were actively using the survey results demonstrating the success of the survey in regards to increasing attention on Framework data. Further responses indicated that some state coordinators had not reached the point of using the response data, but anticipated application of it to their Framework-related work. An important signal of the value of the data

came from a question which identified whether the state coordinators would commit some of their own time to another, similar survey. More than three-quarters of the state coordinators indicated a willingness to commit a week or more of their time to such a process. 54% of state coordinators responded that they would be willing to invest seven days per year into updating the Framework Survey Data. 22% stated twelve days per year and 24% reported one day per year.

### **Conclusions :**

The ultimate success of the survey will not be achieved until it provides coordinators – at either the federal or state level – with a tool to guide their efforts. Still, this report offers a number of findings that can already guide policies and investments in data development. This survey shows that there is a significant foundation for an intensified Framework effort. It also demonstrates that there are any number of weak spots that will need to be patched before the framework can support larger NSDI projects. These weak spots are both thematic and geographic. The survey also shows that varying levels of data quality exist, and this could pose additional problems. Despite these potential problems, the framework appears to be a viable concept with an enormous amount of data available to support its' development.

Of greater concern for the FGDC may be the critical issue of whether the data can be available for nationwide sharing or whether there are currently too many limitations to this initiative. The survey found that many data producing organizations had limited access policies. Additionally, very few organizations were maintaining metadata and, of those that were, few were certain that their data was compliant with the FGDC's Content Standard for Digital Geospatial Metadata. Like the previously mentioned issues this can be addressed, but it needs to continue to receive a relatively high level of attention.

Perhaps the single greatest contribution that the survey project has demonstrated is the difficulty and importance of monitoring the status of data development in the GIS community. Much is being done, little is known, and far too many decisions are being made with inadequate assessments of the nation's digital terrain.

## REFERENCES

- Babbie, Earl. 2000. *The Practice of Social Research*, 9<sup>th</sup> edition. Belmont, CA : Wadsworth Thomson Learning,
- Brown, P. M., and D. D. Moyer (Eds.). 1989. *Multipurpose Land Information Systems: The Guidebook*. Washington, D.C.: Federal Geodetic Control Committee.
- Shutt, Russell K. 1996. *Investigating the Social World: The Process and Practice of Research*. Thousand Oaks, CA: Pine Forge Press:
- Somers, R. 1999. "Framework Data Survey Preliminary Report," Supplement to *Geo Info Systems*.
- Tulloch, D. L. 2000. "Looking Back and Looking Ahead: The State of Framework Data," Proceedings of *URISA 2000*, 427-36 (CD-ROM).
- Tulloch, D. L., and J. Fuld. 2000. "County-level Production of Framework Data: Pieces of a National Spatial Data Infrastructure?" *Journal of the Urban and Regional Information Systems Association*. (Under Review)
- Tulloch, D. L., and M. Robinson. 2000. "A Progress Report on a U.S. National Survey of Geospatial Framework Data," *The Journal of Government Information*, 27: 285-298.