ORGANIZATIONAL NEEDS FOR MANAGING AND PRESERVING GEOSPATIAL DATA AND RELATED ELECTRONIC RECORDS

R. R. Downs[1] and R. S. Chen[2]

[1] *Center for International Earth Science Information Network (CIESIN), Columbia University, 61 Route 9W, Palisades, NY 10964-8000 USA. Email:* rdowns@ciesin.columbia.edu

[2] *Center for International Earth Science Information Network (CIESIN), Columbia University, 61 Route 9W, Palisades, NY 10964-8000 USA. Email:* bchen@ciesin.columbia.edu

ABSTRACT

Government agencies and other organizations are required to manage and preserve records that they create and use to facilitate future access and reuse. The increasing use of geospatial data and related electronic records presents new challenges for these organizations, which have relied on traditional practices for managing and preserving records in printed form. This article reports on an investigation of current and future needs for managing and preserving geospatial electronic records on the part of local- and state-level organizations in the New York City metropolitan region. It introduces the study and describes organizational needs observed, including needs for organizational coordination and inter-organizational cooperation throughout the entire data lifecycle.

Keywords: Geospatial Data, Data Management, Digital Government, Digital Archive, Digital Preservation

1   INTRODUCTION

Government agencies are required to manage and preserve information as official records to enable reliable reuse of this information in the future. Similarly, research institutions and other organizations need to manage data and other scientific and technical information (STI) to preserve these valuable information assets for future access and use (Anderson, 2004). Consistent with recent trends in the use of geographical information systems (GIS) and remote sensing technologies, government agencies and other organizations have become increasingly dependent on geospatial data and related information in digital form (Nelson, Beaverson, & Krumm,, 1998; Maguire, 2003). Geospatial data are used to create data products and maps, to support decision making and perform services, and to serve constituents and customers. In addition to their use by government agencies, decision makers and the public, GIS and geospatial data provide valuable research and teaching resources for faculty who work in various disciplines (Borgman, Smart, Millwood, Finley, Champeny, Gilliland & Leazer , 2005).

Conventions and practices for managing and preserving printed information as official records have evolved over the centuries, enabling organizations to adopt current practices that have been developed and tested within various organizational settings. In contrast, the management and preservation of digital information have a much shorter history of practice and exhibit a high level of technological dependence on evolving hardware and software. The complexity of preserving digital information presents challenges for any organization (Hart & Liu, 2003).

Printing digital documents and managing them as printed records offers organizations a way to continue traditional records management practices, possibly delaying an investment in electronic records management until practices and conventions have matured. However, much information, including geospatial data and related STI that is associated and often integrated with geospatial data, is increasingly difficult to represent completely in printed form. Print technology is a fixed medium of communication and lacks the enhanced functionality offered by the software that is used to create, modify, and present information in digital form. In particular, GIS and related technologies such as Global Positioning System

(GPS) hardware and software offer new capabilities for creating, assembling, integrating, analyzing, visualizing, and managing geospatial data. Printing all possible variations in outputs from such systems would be difficult if not impossible. Management of the geospatial data in electronic form is therefore essential if these data are to be preserved for future use.

The need to manage and preserve geospatial data and related electronic records for future access and use poses new challenges for digital government initiatives and for organizations that create and use geospatial data, especially those with limited resources. Research on digital preservation not only must address a range of technical issues but also "encompasses organisational, legal, cultural, social and financial dimensions" (Hedstrom, Ross, Ashley, Christensen-Dalsgaard, Duff, Gladney, Huc, Kenney, Moore, & Neuhold, 2003). Organizational issues are as important as technical issues for preserving access to digital assets (Carpenter, 2005). Practices in institutions, including government agencies and other organizations, can be informed by research that addresses the organizational challenges for the preservation and management of digital assets.

In this paper, we report on organizational needs for managing and preserving geospatial electronic records (GERs) on the part of local- and state-level organizations in the New York City metropolitan region. This study is part of a larger project to assess how government agencies and other organizations can address challenges for managing and preserving geospatial electronic records, described by Chen and Downs (2004). Other project outputs include a Guide for Managing Geospatial Electronic Records (2005) and a Data Model for Managing and Preserving Geospatial Electronic Records (2005). Subsequent papers will address additional aspects of the project.

2   STUDY DESIGN

Our study employed a qualitative approach to investigate the needs for managing and preserving geospatial data and related electronic records. Professionals representing a variety of disciplines, levels of responsibility, and types of organizations were recruited to observe a cross-section of perspectives on organizational practices for managing and preserving these data. Interviews were conducted to assess the practices of a diverse group of professionals who use and manage geospatial data in their organizations. Analysis of the interview responses revealed the current and future needs of these professionals for managing and preserving geospatial data within their organizations.

2.1   Participants

In this study, we sought to reflect the range of professionals and organizations that acquire, use, and manage geospatial data in a specific region, the four states of New York, New Jersey, Connecticut, and Rhode Island. As an initial step, we established a project Advisory Board consisting of fourteen professionals who share an interest in improving practices for managing and preserving geospatial data and related electronic records. Members included geospatial data librarians and coordinators, archivists and electronic records managers, GIS analysts and directors, and planning and policy analysts. They serve in various roles within state, county, and local government agencies and other organizations in the four states. The Advisory Board members included eight males (57%) and six females (43%), including several individuals suggested by the National Historical Publications and Records Commission (NHPRC) staff.

Participants in the study were recruited from a diverse pool of candidates nominated by members of the Advisory Board. Individual members helped us to create a pool of ninety-five nominated candidates, who represented a diverse set of professionals employed by government agencies and other organizations in all aspects of managing and preserving GERs.

The nominated interview candidates were all professionals who worked primarily in locations within the northeastern region of the United States, including New York State (49%), Connecticut (24%), New Jersey (19%), and Rhode Island (3%), but also other states (4%). The pool of candidates consisted primarily of males (75%), compared with females (25%). Table 1 presents the distribution of the nominated interview

candidates by the levels of government and the types of organizations in which the individuals were employed.

Table 1. Distribution of Nominated Candidates by Organizational Affiliation

Organizational Affiliation	Percent
Federal Government	5%
State Government	23%
County Government	21%
Municipal Government	20%
Educational Institution	20%
Non-Profit Organization	8%
Commercial Organization	2%
Total	100%

2.2 Interviews and Analysis

An interview guide containing an informed consent script and a semi-structured interview questionnaire was developed in consultation with the Advisory Board. The interview questionnaire contained twenty-two questions organized within the following categories ordered to elicit descriptions from each participant regarding their own practices, the practices of their unit or department, and the integration of practices within their organization for managing and preserving the geospatial data and related electronic records:

- Data acquisition, use, and management
- Current records management infrastructure
- Metadata
- Preserving access
- Barriers and enablers
- New tools and resources
- Other issues

Each question category consisted of open-ended questions to provide participants with an opportunity to describe aspects of their practices for managing and preserving geospatial data and related electronic records and to elaborate on any issues mentioned, whether related to resources and procedures used or managerial or organizational concerns. The interview questionnaire is presented in section 8.1 of the Appendices.

An investigator called the telephone number provided for each of the candidates to request participation in the study. Interviews were scheduled with those candidates who agreed to be interviewed for up to an hour. For consistency, the same investigator conducted all of the telephone interviews, which were completed between January 2004 and February 2005.

All of the interviews were conducted by telephone and usually lasted between forty minutes and one hour. During each telephone interview, the interviewer used the interview guide to ask questions within each category and to write field notes capturing the responses. The interviewer encouraged participants to elaborate on their responses by asking open-ended follow-up questions. The interviewer subsequently annotated the field notes immediately following each interview to insure legibility and completeness. In addition, the interviews were recorded on audiotape, with the exception of those interviews and interview segments that occurred when participants requested the interviewer to stop recording.

Initially, the field notes from the responses to all of the questions in each interview were read to identify issues from each participant that represented needs for managing and preserving geospatial data. The tape recordings were employed to further annotate field notes that required additional clarification during the analysis of the interview responses. Identified issues were captured and organized into initial categories representing common themes as they developed. Subsequently, the responses from each of the interviews

were read within the context of each question to identify additional issues among the responses of the participants. As issues were identified and compared with previously identified issues, they were used to establish new themes or categorized within subcategories of the emerging themes.

Analysis of the interview responses identified categories of organizational needs of government agencies and other organizations for managing and preserving the geospatial data and related electronic records that they create and use to support analysis, decision-making, or the creation of new products to serve their customers. Within these categories, patterns emerged that reveal the organizational needs of the study participants. Resulting themes and issues pertinent to organizational needs are described in the following sections.

3   RESULTS AND DISCUSSION

Thirty-one professionals representing twenty-seven organizations were successfully interviewed. Each participant held a position within their organization with some responsibilities for managing geospatial data or other records. A wide range of professional responsibilities were represented among the interviewed participants in terms of their professional roles, their organizational authority, and the levels of government and other types of organizations in which they were employed.

The participants included geographical data analysts and developers of geographical data products (26%), geographical data librarians (10%), archivists and records managers (13%), information systems and technology directors (10%), geospatial data coordinators or managers (32%), and directors of geospatial or planning departments (10%). Most of the participants worked in New York State (61%) and New Jersey (19%). The remainder of the participants worked in Connecticut (6%), Rhode Island (6%), and other states (6%). Twenty-six (84%) of the participants were male and five (16%) of the participants were female. Table 2 presents the proportional representation of the types of organizations that employed the study participants.

Table 2. Distribution of Study Participants by Organizational Affiliation

Organizational Affiliation	Percent
Federal Government	3%
State Government	42%
County Government	16%
Municipal Government	13%
Educational Institution	13%
Non-Profit Organization	13%
Total	100%

The interviews revealed that the participants are seasoned and committed professionals who are using limited resources within a technologically complex domain to meet the needs of various constituents. The participants considered the need to improve geospatial data management practices as a high priority. Many participants expressed an interest in the study and felt that research is needed to improve practices for managing geospatial data. In a few cases, participants who were initially unsure of their potential contribution to the study disclosed that their practices were not ideal for managing or preserving their geospatial data, even though the necessity for managing these resources is recognized within the organization. Similarly, participants often articulated the urgency for improving their practices and stated that the importance of studying these issues justified the use of their time to participate in the study. Generally, participants recognized that managing and preserving geospatial data and related electronic records presents critical challenges for their organizations. The study identified two categories of organizational needs that influence the management and preservation of geospatial data and related electronic records by government agencies and other organizations. These observed categories are:

- Coordination of the Data Lifecycle

- Inter-Organizational Cooperation

These categories represent non-trivial challenges and opportunities for managing geospatial data and related electronic records as valuable organizational assets that represent considerable investments by the organizations involved in creating and using these resources. The following sections describe the challenges observed within each of these categories.

3.1 Coordination of the Data Lifecycle

Interviews of many of the participants revealed that disparate units within a single organization engage in similar activities to meet shared organizational objectives for managing geospatial data. However, several professionals reported that without overall coordination of these activities, the members of the relevant workgroups often found themselves working inefficiently and compromising shared organizational efforts throughout the data lifecycle. Organizational activities that were identified by the participants as needing coordination throughout the data lifecycle are:

- Organizational commitment to the geospatial data management lifecycle
- Data acquisition and development;
- Quality assurance, documentation, and record-keeping; and
- Archiving and preservation

Observed needs for organizational coordination of the data lifecycle are described for each of these areas in the following sections.

3.1.1 Organizational Commitment to the Geospatial Data Management Lifecycle

Many participants reported that their own organization needs to recognize the value of managing geospatial data and the importance of a coordinated organizational effort to ensure that geospatial data are being managed effectively throughout the entire data lifecycle. Several participants described a similar pattern of events that successfully led to improvements for managing geospatial data: first, recognition for geospatial data development by constituents; second, recognition by top management of the importance for acquiring, developing, and managing geospatial data and for establishing priorities for managing these data; and third, allocation of resources to manage these data effectively. This commonly repeated sequence of events enabled the individuals responsible for geospatial data management and operations to improve their practices.

The participants strongly emphasized that the management of geospatial data needs to be recognized as important by top management and supported by managerial and professional staff across the organization in a coordinated manner. Simply assigning the management of geospatial data to a unit with limited resources and competing priorities can lead to compromises that degrade the quality of the data management effort. Taking an approach that commits the overall organization and its resources to the management of geospatial data enables members of the organization, regardless of their position within the management structure, to contribute effectively to managing these data from acquisition through subsequent development and preservation efforts. It also provides those involved in geospatial data management with an opportunity to recognize the value of their contributions to the organization, in contrast to those who described frustrations experienced in the absence of organizational coordination.

More than a few participants, including geographical unit directors and managers, recognized that leadership is needed to initiate and sustain organizational coordination for managing geospatial data. Participants serving in roles as organizational leaders often commented that they need to be able to justify the commitment of resources to managing geospatial data. A few managers who reflected on the improvements achieved within their units reported that an integral aspect for leading such an effort is recognition, by organizational stakeholders, of the value of managing geospatial data. They believed that recognition by stakeholders, including those within the organization and those served by the organization,

enabled them to develop policies and follow through with the resources to implement their plans for managing geospatial data in an effective and coordinated manner.

Participants serving in roles as geospatial data librarians and coordinators emphasized that organizational leaders need to demonstrate this commitment to the entire organization by establishing policies to manage geospatial data and related electronic records as official records of the organization. Combined with stakeholder recognition for the value of managing geospatial data, these policies and plans provided justification to authorize the use of organizational resources to manage geospatial data effectively. Recommendations from a workshop on managing and preserving geospatial electronic records included demonstrating the benefits of geospatial data management to assist leaders who need to justify the allocation of organizational resources for managing and preserving these data (Downs & Chen, 2005).

The leaders of the organizations that have committed to managing geospatial data also realized that these responsibilities begin early in the data lifecycle, before data acquisition and development have been initiated. Specific individuals were given the responsibility and the authority at an early stage to coordinate with others during the data lifecycle and to plan for and use the necessary resources to effectively acquire, manage, and preserve their geospatial data and related electronic records.

3.1.2   Data Acquisition and Development

Types of geospatial data acquisition described by participants included purchasing or acquiring data without cost, developing new data with their own staff or in collaboration with others, procuring the services of others to develop data, and developing data by combining aspects of existing data that had been developed by their own organization or acquired from others.

Interestingly, many participants reported acquiring geospatial data from the same data providers. These included government agencies, data centers, data clearinghouses, and commercial data providers. According to several participants, they and others within their organization routinely relied on the documentation that was provided with the data to determine whether to acquire and use the data. The documentation of interest includes documentation of previous reviews and appraisal of a dataset. Participants also reported that the data provided by some known sources are appropriately documented, facilitating efficient appraisal. By acquiring their geospatial data from these known sources, the appraisal decision could focus primarily on the acquisition of data collected with a particular instrument or methodology, within a specific time period, and for a particular location.

Some participants elaborated on how they repeatedly used the same data providers to obtain their data and that acquiring geospatial data from these known sources reduced their geospatial appraisal costs. Although the level of trust was not the same for each source, the trustworthiness of a data provider influenced the geospatial data appraisal process when acquiring data from a particular source. Participants determined their level of trust for a data provider primarily based on their own experiences and the experiences of others in their organization. However, they also based their level of trust on reports received from other members of the geospatial data community regarding a data provider and its collections. For many organizations, the reputation of a particular data provider within the community also influenced the amount of trust for the data provider. Nevertheless, defining the current and future designated community of users presents challenges to data providers for documenting, supporting, and determining appropriate formats for disseminating and preserving geospatial data (Parsons and Duerr, 2005).

Organizations need to allocate sufficient staff and resources to effectively conduct the geospatial data management effort. A few participants were frustrated that they were expected to improve geospatial data management practices in the absence of additional resources or personnel assignments. Some reported that they would not let the lack of resources frustrate them, and accomplish what they could with the resources received. Simply assigning geospatial data creators and developers with additional tasks for managing and preserving these data does not necessarily contribute to improved data management and preservation. Many participants reported that they requested additional staff and suggested that individuals from various functional areas of the organization need to be reassigned, in some cases, to focus the organization's resources on managing geospatial and related electronic records.

Many participants reported experiencing the challenges of trying to develop geospatial data without adequate resources. Geographers, GIS analysts, and geospatial data developers described their commitment to ensuring that the data that they acquire and develop meets the requirements for the project. When short deadlines and limited resources impose constraints on a project, these skilled professionals do not and should not compromise project goals that ensure the quality of the data acquisition and development effort. However, without sufficient staff and resources assigned to the project, participants reported that efforts to achieve organizational goals for data management and quality reflected these deficiencies.

Interestingly, those reporting from organizations that have assigned information technology (IT) professionals, archivists, and records management professionals to work with the geospatial data professionals on these projects described coordinated acquisition, accession, management, development, and preservation of these data in a controlled manner. By bringing this expertise to bear, these organizations improved the management of data developed for earlier projects and ensured that later projects also could benefit by reusing the data and the methods applied earlier to develop these data. Similarly, Downs and Chen (2003) describe how the establishment of cross-disciplinary teams can contribute to the development of interdisciplinary data products and to the provision of services that support their use.

Participants in a few organizations described how liaisons are assigned from units involved in geospatial data development to work with their counterparts employed in centralized IT, archival, and records management units. GIS analysts who described how aspects of data management and preservation were being completed by IT and records management staff also described how these activities were coordinated by individuals who served as liaisons to those units. In contrast, a few GIS analysts and data product developers who worked in isolation lamented that they had not addressed data preservation and that they would like to improve their understanding on how to store and format their data for future use. Similarly, others described assorted storage media, containing potentially useful geospatial data, which could no longer be accessed with their current technology.

Data that have been acquired or developed for previous programs also have potential value for future programs. More than once, members of these organizations described the disappointment when they learned that they had to duplicate their data development efforts from earlier projects when conducting similar projects just a few years later. Staff turnover and incomplete understanding of earlier assumptions and decisions forced them to reproduce results that they might well have used otherwise.

It is also critical to budget appropriately to maintain data of continuing value that has been initially acquired or developed for previous programs. The data that had continuing budgets were given priority and received the attention. Budgets need to be allocated to maintain data management operations for geospatial data that has been appraised as possessing continuing value for the organization.

A few participants also reported that errors can occur when data that are acquired or developed are not sufficiently identified to differentiate them from other data being managed. Those who reported learning costly lessons suggested that providing access to the most current version of a dataset fosters use of the correct version. They also recommended clearly labeling the storage media containing each dataset with the version number and the date of release, as well as the title, to differentiate it from previous versions that may have been developed with the same title during the same time period. Paskin (2005) describes how digital Object Identifiers (DOI) can be applied to scientific data to establish unique and persistent identifiers for accessing scientific data.

3.1.3   Quality Assurance, Documentation, and Record-Keeping

According to participants who considered their organizations successful at managing geospatial data, appraisal takes place throughout the entire data lifecycle. In addition to formal appraisal practices to select data for acquisition and development, more than one participant described the necessity of conducting a "layer by layer" committee review of datasets for dissemination and archiving.

Participants also described the necessity for engaging in electronic record-keeping practices. For example, many reported that their organizations need to track data development and dissemination efforts being conducted for current and past projects. This documentation includes identifying the specific collection, sub-collection, series, project, or project phase that sponsored the creation, acquisition, or dissemination of a particular data set as well as identifying the individuals involved and their roles. It also includes documenting any historical or intermediate versions of data completed during data development, the completion dates, and the original data set or data sets that contributed to the development of data derived from multiple sources or completed during successive projects. Participants suggested that keeping information about the project and the collection is necessary for managing the data and understanding the current status and the decisions made about the data throughout its lifecycle.

Because the effort to manage data and information can be taxing on organizations and projects, participants suggested identifying a common minimum set of metadata elements to be required. Participants in some organizations reported that a set of minimum requirements apply to all data managed by the organization, regardless of the project. In addition, they reported that each project has its own requirements to meet, which may be more demanding than the requirement for other projects.

Another consideration for the data lifecycle that some participants suggested is revising the documentation on older datasets to be consistent with adopted standards so that the quality of all of the data that is being managed is consistent. For example, when some organizations began storing newly acquired metadata in an XML-based relational database, they converted the metadata describing older datasets to be compliant with the new standards. Regardless of how the data is stored, several participants stated that they preferred to receive discovery metadata in text format along with data so that they were not dependent on software, which could become antiquated over time, to use the metadata.

Participants often described simple methods for organizing their data. Several participants mentioned that they maintained thematically named folders and file naming conventions to organize their data and to ensure that the data could be managed effectively. Similarly, a few reported using file naming schemes to identify aspects of derivative products created from each data source.

The record-keeping practices for geospatial data in many of the small agencies observed were reported as not being adequate or consistent for the data being managed by a particular organization. Participants who managed small departments in larger organizations described similar circumstances. Often, the inadequate record-keeping practices were perceived as resulting in unusable data. Also, inconsistencies in record-keeping often resulted in duplication of efforts, which wasted scarce resources and which participants described as frustrating.

Some participants explained that organizational coordination is also needed to ensure that documentation can foster capabilities for creating and using derivative data products created from one or more data resources. They said that data development activities need to be coordinated to foster sharing of information and to effectively document techniques that are employed while developing data products, including any assumptions, problems encountered, and conclusions drawn, as well as describing methods used for integrating datasets. Bose and Frew described similar symptoms when identifying the challenges for describing the processing of geospatial data products, noting that "no definitive method, standard, or mandate exists for preserving, providing, or communicating the lineage of computational results" (p 5, 2005).

Many participants also reported that sharing and documenting lessons learned during data development contribute to organizational learning. They suggested that knowledge about the development of the data is important when reusing the data. In addition, a couple of participants reported that such knowledge enabled others to use and build on such techniques when engaging in similar data development activities.

Many of the participants reported that they have established custom databases to meet the operational needs for managing geospatial data projects being completed by an individual, group, or organizational unit. Often, the degradation of the utility of a custom database for dataset management was due to inadequate descriptions of the data necessary to support ongoing management and preservation purposes. Several

reported that their custom databases for managing geospatial data were designed primarily to support the operational aspects of projects. A few participants mentioned that the eventual use of these databases later in the data life cycle was not envisioned or considered during the initial design, but they enhanced their databases as their needs evolved.

The custom databases include information about geospatial data holdings, but do not necessarily contain the structural, administrative, and preservation information about individual datasets for managing and preserving them beyond the life of a particular project. The collection of some relevant information for managing geospatial data in such databases demonstrates the need to include relevant collection, series, project or other classification information in record-keeping systems used for managing and preserving geospatial data.

Some short-lived successes were reported for managing data while the data were being used for current projects and while those working with the data retained some familiarity with the provenance of the data. One participant described how the capabilities for using a custom database to manage a particular geospatial dataset degraded after the project employing the dataset had been completed. A few others described how shifting priorities and the lack of a robust database design contributed to the loss of capabilities to record the provenance and preservation information needed to effectively manage and preserve the data.

Frequently, limited descriptions of geospatial datasets resulted from staff turnover in participants' organizations and from new priorities diverting the attention of responsible individuals to data that were being developed or acquired for the next project. While staff turnover can negatively affect the quality of operations in organizations, a few participants described how vigilant record-keeping can reduce the impact on processes and on products produced.

3.1.4   Archiving and Preservation

Most of the interviewed participants use the term "preservation" in a consistent manner. In contrast, the term "archive" is used with various meanings. Not surprisingly, almost all of the participants expressed concerns for preventing the unintentional loss of data.

However, although many participants could describe an organizational records management program, very few participants regarded their own geospatial data management efforts as part of such a records management program. On the contrary, several participants considered these efforts as separate and quite different functions. Furthermore, when asked to consider the needs for managing geospatial data as electronic records, participants often described similar challenges confronting their organization for managing and preserving both geospatial data and other electronic records. The National Science Board (2005) also has recognized challenges for preserving scientific data and has produced recommendations for the scientific community.

Participants whose organizations had instituted electronic records management reported storing their geospatial data in a centralized repository, suggesting that managing a centralized repository simplified management of the data. Similarly, these participants reported that their organizations designated one work group to manage the archival facility where the geospatial data and other electronic records were maintained for the organization. Each group that acquired or developed data provided the data to the archival facility. Giving one group the authority to manage and preserve geospatial data along with other types of data ensured that the responsibility was assigned to those who could focus on managing the data.

A few participants described the efforts of their organization to implement a centralized digital repository system to manage all of the data and the metadata contained in organizational collections. Those who described digital repository development efforts also described challenges associated with these efforts, stating that an organizational commitment to developing and using a digital repository is necessary to meet such challenges. One participant described building a trusted institutional repository to enable everyone in the organization to deposit data for preservation and access. Recently, attributes have been described for repositories to be trusted with the preservation of digital objects (RGL, 2002).

Participants described how they developed formal specifications for the preservation of geospatial data that improved the data stewardship practices within their organizations. Some participants who considered their organizational efforts successful for managing and preserving data described having policies and procedures to prevent the loss of data and to ensure the preservation of data that was identified as important. A few participants from these organizations stated that such policies included details to ensure continuing data stewardship, such as specifications for routinely archiving data on compact disk (CD) and digital video disk (DVD) media and storing one copy off site. Some procedures that were described by participants also specified frequent delivering of system backup tapes and copies of data to both onsite and offsite storage locations in order to immediately protect the data from potential hazards that could affect one site. Organizations need to determine the appropriate distance from their primary archive when identifying a location for storing copies offsite (Entlich, 2005).

In a few organizations, participants found it necessary to carefully control versions of datasets. When new versions were created, the titles of the newer versions often were the same as the older versions, which often caused confusion. Participants described this as a common problem that occurred in the absence of adequate record-keeping practices. More than one participant described how their organization addressed the issue by carefully labeling media and archiving older versions of datasets offline, separate from the newer versions that had been acquired or developed.

Most participants recognized that documentation is a key aspect of preserving data. A few participants described how they regularly document the techniques and methods developed and used for integrating data. On the contrary, others described cases in which data could not be used because methodological assumptions and conclusions that were made during the data integration process were not adequately documented and that those who were familiar with the process since left the organization. Those participants who considered their organization's archiving and preservation practices to be successful also had established minimum documentation requirements for geospatial data. For example, they described using templates to guide practices for creating documentation for their geospatial data and to ensure completeness in their documentation efforts. The Consultative Committee for Space Data Systems (2002) has produced recommendations for archival information packages (AIP) to describe information for long-term preservation in an Open Archival Information System (OAIS).

Participants within several organizations reported that education and training are critical for improving archiving and preservation efforts. Some participants suggested that their organization's data management efforts could be enhanced to include training and cross-training programs that equip various staff members with the range of skills necessary to manage geospatial data. Many participants described how official organizational records are being created, stored, and used in digital formats, but archival and records management practices continue to be limited to the traditional methods of managing printed records only.

Professionals in organizations that have begun exposing archival and records management professionals to practices for managing electronic records reported that this exposure helped them to become familiar with current archival methods and records management techniques, which they have begun to apply to manage geospatial data as electronic records. Similarly, organizations that have exposed their geospatial data analysts and product developers to archival and records management practices have educated these professionals so that they can help prepare geospatial data for archiving and contribute to records management and preservation activities. Furthermore, the archivists and records managers who were exposed to practices for creating and using geospatial data reported that this exposure has increased their sensitivity to the functional aspects of these data that must be preserved digitally. Several participants indicated that they need additional expertise for archiving and that guidelines and standards for archiving geospatial electronic records would be helpful for themselves and for members of their staff.

Many of the observed small agencies and small departments that use geospatial data are not managing these data as electronic records, resulting in the loss of legacy data and related information that are valuable organizational assets. In many cases, individuals responsible for records management within these organizations have traditionally worked with physical, printed records, such as maps. However, some of the traditional archivists and records managers interviewed have begun to gain familiarity with the issues and

demands of managing electronic records. Nevertheless, pervasive issues such as evolving technology, proprietary software requirements, and the multiplicity of data formats continue to increase the complexity for managing geospatial data and will continue to present challenges for records managers. These concerns are consistent with findings by Borgman et al. (2005) on faculty use of geospatial data obtained from others.

3.2   Inter-Organizational Cooperation

Many participants expressed a need to learn how other organizations are managing their geospatial data, indicating that it is inefficient to try to develop practices that other organizations already have developed. Several participants also said that they are willing to share what they have learned with individuals in other organizations and that the feedback received from others often leads to improvements. Overwhelmingly, participants considered cooperating with other organizations to be a success factor for improving their practices.

Many participants described the ways in which they cooperate with other organizations that result in mutually beneficial improvements in their efforts to manage geospatial data. Such inter-organizational cooperation includes collaborating on projects; sharing skills, procedures, specifications, training and professional development opportunities; and identifying contacts within other organizations to attain specific objectives or resolve a particular problem. Participants' statements regarding inter-organization cooperation are consistent with observations that organizations can improve practices for managing and preserving geospatial data by cooperating with other organizations (Downs & Chen, 2003). By cooperating with other organizations, the staff within each cooperating organization can share their knowledge and assist each other in attaining common goals.

The following categories were observed for improving inter-organizational cooperation:

- Partnerships and Inter-Organizational Cooperation
- Community Resources

The needs for improving inter-organizational cooperation are described for each of these categories.

3.2.1   Partnerships and Inter-Organizational Cooperation

Most participants described how their organizations engage in efforts to acquire or share data with other organizations. Since geospatial data can be costly to create or acquire, many participants found it beneficial to enter into cooperative relationships with other organizations.

Participants described many needs of their organizations to respect any restrictions for access and use, citation requirements, and intellectual property (IP) rights of geospatial data acquired from other organizations. Some participants discussed how negotiations with other organizations can become complicated and time consuming when establishing data sharing agreements. A few participants described how they reduced this complexity by developing and maintaining written agreements for data sharing. By developing written agreements that could be used with various partners, data managers did not have to initiate new agreements for each negotiation to acquire or obtain a dataset. A reusable template could be maintained and used as a "stock" agreement that is improved as terms that can serve all parties are recognized for inclusion in the template and added to the agreement for use in future negotiations.

Several participants reported that their organization identifies and records access restrictions, citation requirements, and IP rights during acquisition of the data, rather than after using the data or sharing the data with others. Similarly, some participants described that it is necessary to specify any sensitivity and security concerns when developing data sharing agreements with other entities. Raising and documenting the data-specific security and sensitivity issues in written agreements pertaining to the dataset reduced the ambiguity that otherwise could become problematic and time-consuming when questions are asked about such issues in the future. Sieber (2005) also describes various ethical issues and related concerns that have risen from practices of sharing scientific data.

In a few cases, participants reported that their organization has established policies to not disseminate data acquired from other organizations and to routinely restrict such data to internal use only. Some participants who share data with other organizations reported that they specify usage constraints for such data that include a statement absolving the creators and the data provider from any liability. Many of these needs for data sharing are consistent with the confidentiality and ethical issues of using geospatial data described by Golden, Downs, and Davis-Packard (2005) and with the observations of Borgman *et al.* (2005), who described difficulties that intellectual property rights pose for sharing geospatial data among research community members.

Due to the effort required to appraise data, one participant felt that it would be valuable for partnering organizations to establish a peer-review council to review data to be released for public use or included in clearinghouses. The participant suggested that each partnering organization in the group could contribute staff time to the joint appraisal effort and reduce the redundant appraisal of data that is taking place within each organization.

Almost all participants described how they preferred receiving all data documentation in digital form and that this was the expected practice and the norm for most organizations that cooperate with others. One participant said that they also require all acquisition contract documents to be written in digital form, in an ASCII-based standard format such as SGML or XML. Requiring contract documents in digital form enabled such documents to "stay with the data," so that these contracts could be managed with the data and not become separated.

A few participants reported that they are forming web services partnerships to provide access to and integrate with data served by others. Their comments suggested that such partnerships will result in improved capabilities, but such projects do require mutual sharing and much time dedicated to these efforts.

Several participants stated that organizations need to share lessons learned on implementing digital repositories and similar technologies. They reported that they do not have the time themselves and that others also do not have extra time to adopt new technologies that require excessive time to implement. A few suggested that receiving information about similar efforts by others could help them and that they would contribute information about their experiences to a library of case examples, best practices, and recommended technologies for improving geospatial data storage and delivery.

3.2.2   Community Resources

Interestingly, many of the participants described how they use collections of accessible geospatial data that serve as community resources for the organizations within the geospatial community. These services include publicly accessible clearinghouses of geospatial data, online catalogs and registries of geospatial data available from those who host them, and data centers that are sponsored by government and academic institutions. Those who used these digital libraries and similar community resources acclaimed the benefits of using these services. One participant stated that she would not be able to do her job without the availability of the web-based catalogs of free, peer-reviewed, online accessible geospatial data.

A few participants suggested improvements that would increase their capabilities for acquiring, using, developing, and managing geospatial data. For example, one participant recommended that these community service providers offer Web-browser access to facilitate discovery of data layers, conduct GIS analysis, and engage in interactive mapping. Others suggested the need to sustain these types of community services so that those who use them could rely on these services as persistent archives and reduce their own organizational needs to manage the data.

Although the use of community tools was currently limited among the participants interviewed, many offered suggestions for establishing shared community tools when asked to recommend resources that their organization would find useful. Several participants suggested maintaining accessible documentation on best practices of community members so that each organization could contribute and obtain documentation on geospatial data management that would be stored within a distributed and easily searchable database.

One participant suggested that such a database would be valuable if it enabled community members to share lessons learned on implementing digital repositories.

Even those participants who described their organization's geospatial data management as being successful reported that meeting the challenges of preserving geospatial data will require efforts to pool resources and engage in collaborative inter-organizational efforts. Those participants who described their experiences sharing resources with other organizations found that efficiencies were attained as the result of such collaborations and that their learning was heightened. The practices of open source software development projects also demonstrate how sharing information online can be beneficial for contributing members of the community (Lakhani & von Hippel, 2003).

Others described their willingness to contribute to community resources. For example, one participant who represented a government agency stated that the time to serve on a peer-review council to review data could be justified if the data were to be released for public use or included in clearinghouses of freely accessible data.

The observations on community resources contribute to our understanding of the benefits of collaboration that can be gained through cooperative development and management of geospatial data as described by Downs and Chen (2003). However, these observations also indicate that additional research is needed on pooling resources with collaborators in other organizations to increase capabilities for geospatial data acquisition, development, management, and preservation.

4 LIMITATIONS

The small sample size of thirty-one interviewed participants limits the generalizability of the findings of this study. Additional studies of professionals who manage geospatial data could help to generalize the findings to the larger population of professional geospatial data managers. Also, the recruitment of participants interviewed for this study limits the generalizability of the categories of professionals and organizations that manage geospatial data. Participants were recruited for the study as a result of their recognition within the communities of professional archivists, geospatial data managers, and geographical analysts. Although the participants of this study reflected a diverse group of professionals, future studies may need to identify participants representing other professions that engage in geospatial data management practices. The participants interviewed in this study primarily represent organizations and agencies located within the New York metropolitan region of the United States and may reflect regional practices, which might differ from practices in other regions of the U. S. and in other countries. Observing practices of professionals in other regions of the U. S. and in other countries could identify aspects of geospatial data management practices that are more generalizable nationally or worldwide.

In addition, the interviews obtained self-reported statements from participants who answered general questions about their geospatial data management practices. Even though participants received assurances of anonymity, the objectivity of interview responses describing self-reported practices cannot be ensured. Additional studies are needed that use other observational techniques, such as direct observation, to validate our results. Also, in-depth interviews of specific techniques and processes used by professionals to manage geospatial data would contribute to understanding of geospatial data management practices.

5 CONCLUSIONS

This study observed the practices for managing geospatial data and related electronic records of thirty-one professionals, representing twenty-seven organizations primarily from the New York metropolitan area of the United States. Analysis of the practices of the interviewed participants has revealed patterns of needs to be addressed for organizations to improve practices for managing and preserving geospatial data for continuing and future access and reuse. Although some of the observed organizations that have begun addressing these needs have improved their data stewardship and electronic records management practices, the professionals within these organizations also recognize the need to continually work with others to meet

these challenges and to improve their practices for managing and preserving geospatial data and related electronic records.

6   ACKNOWLEDGEMENTS

7   REFERENCES

Anderson, W. L. (2004). Some Challenges and Issues in Managing, and Preserving Access to, Long-Lived Collections of Digital Scientific and Technical Data. *Data Science Journal*, *3* (30 December), 191-202. Retrieved 6 June, 2005 from the world wide web: http://www.datasciencejournal.org

Borgman, C. L., Smart, L. J., Millwood, K. A., Finley, J. R., Champeny, L., Gilliland, A. J., & Leazer, G. H. (2005). Comparing Faculty Information Seeking in Teaching and Research: Implications for the Design of Digital Libraries. *Journal of the American Society for Information Science and Technology*, *56*(6), 636-657.

Bose, R. & Frew, J. (2005). Lineage Retrieval for Scientific Data Processing: A Survey. *ACM Computing Surveys*, *37*(1), , 1–28.

Carpenter, L. (2005). Supporting Digital preservation and Asset Management in Institutions. *Ariadne*, *43*, (April). Retrieved 6 June, 2005 from the world wide web: http://www.ariadne.ac.uk/issue43/carpenter/

Center for International Earth Science Information Network. (2005). *Data Model for Managing and Preserving Geospatial Electronic Records*. Retrieved 6 June, 2005 from the *CIESIN* website: http://www.ciesin.columbia.edu/ger

Center for International Earth Science Information Network. (2005). *Guide to Managing Geospatial Electronic Records*. Retrieved 6 June, 2005 from the *CIESIN* website: http://www.ciesin.columbia.edu/ger

Chen, R. S. & Downs, R. R. (2004). Managing and Preserving Geospatial Electronic Records. Presentation to *Extending the Boundaries: A Workshop on Managing and Preserving Geospatial Electronic Records*. Columbia University Lamont Campus, Palisades, NY. Retrieved 6 June, 2005 from the *CIESIN* website: http://www.ciesin.columbia.edu/ger/workshop.html

Consultative Committee for Space Data Systems (2002). Reference Model for an Open Archival Information System (OAIS). Retrieved 6 June, 2005 from the world wide web: http://www.ccsds.org/documents/650x0b1.pdf

Downs, R. R. & Chen, R. S. (2005). Extending the Boundaries: A Workshop on Managing and Preserving Geospatial Electronic Records, Report of the Workshop. Retrieved 6 June, 2005 from the *CIESIN* website: http://www.ciesin.columbia.edu/ger/workshop.html

Downs, R. R. & Chen, R. S. (2003). Cooperative Design, Development, and Management of Interdisciplinary Data to Support the Global Environmental Change Research Community. *Science &*

Technology Libraries, 23(4), 5-19. Co-Published Simultaneously in Virginia Baldwin (Ed.), *Online Ecological and Environmental Data*, Binghamton, NY: Haworth Press

Entlich, R. (2005). Too Close for Comfort? The Case for Off-site Storage. *RLG DigiNews* 9(6). ISSN 1093-5371. Retrieved 6 June, 2005 from the *RLG* website: http://www.rlg.org/preserv/diginews/

Golden, M. L., Downs, R. R., & Davis-Packard, K. (2005). *Confidentiality Issues and Policies Related to the Utilization and Dissemination of Geospatial Data for Public Health Applications*. A Report to: The Public Health Applications of Earth Science Program, National Aeronautics and Space Administration, Science Mission Directorate, Applied Sciences Program. Socioeconomic Data and Applications Center (SEDAC), Center for International Earth Science Information Network (CIESIN), Columbia University, March 2005. Retrieved 6 June, 2005 from the *CIESIN* web: http://www.ciesin.columbia.edu/pdf/SEDAC_ConfidentialityReport.pdf

Hart, P. E. & Liu, Z. (2003). Trust in the preservation of digital information. *Communications of the ACM*, *46*(6), 93-97.

Hedstrom, M., Ross, S., Ashley, K., Christensen-Dalsgaard, B., Duff, W., Gladney, H., Huc, C., Kenney, A. R., Moore, R., & Neuhold, E. (2003). *Invest to Save: Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation*. Retrieved 6 June, 2005 from the *ERPANET* website: http://eprints.erpanet.org/archive/00000048/

Lakhani, K. R. & von Hippel, E. (2003). How Open Source Software Works: "free" user-to-user assistance. *Research Policy*, *32*(6), 923-943.

Maguire, D. (2003). The Impact of Global Change on the Future of National Mapping Agencies. *Cambridge Conference 2003*. Cambridge, UK. Retrieved 6 June, 2005 from the world wide web: http://www.cambridgeconference2003.co.uk/camconf/papers/6-3.pdf

National Science Board (2005). *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. National Science Foundation. Retrieved 6 June, 2005 from the *NSF* website: http://www.nsf.gov/pubs/2005/nsb0540/

Nelson, D. O., Beaverson, S. K., & Krumm, R. J. (1998). The Illinois Natural Resources Geospatial Data Clearinghouse: A Prototype NSDI Clearinghouse Node in Illinois. *Digital Mapping Techniques '98 -- Workshop Proceedings*. U.S. Geological Survey Open-File Report 98-487. Retrieved 6 June, 2005 from the *USGS* website: http://pubs.usgs.gov/of/of98-487/nelson.html

Parsons, M. A. & Duerr, R. (2005). Designating User Communities for Scientific Data: Challenges and Solutions. *Data Science Journal*, *4*(24 August), 31-38. Retrieved 6 June, 2005 from the world wide web: http://www. datasciencejournal.org

Paskin, N. (2005). Digital Object Identifiers for Scientific Data. *Data Science Journal*, *4*(28 April), 12-20. Retrieved 6 June, 2005 from the world wide web: http://www.datasciencejournal.org

Sieber, J. E. (2005). Ethics of Sharing Scientific and Technological Data: A Heuristic for Coping with Complexity & Uncertainty. *Data Science Journal*, *4*(22 December), 165-170. Retrieved 6 June, 2005 from the world wide web: http://www.datasciencejournal.org

RGL (2002) Trusted Digital Repositories: Attributes and Responsibilities. Mountain View, CA: RLG, May 2002. Retrieved 6 June, 2005 from the *RGL* website: http://www.rlg.org/en/pdfs/repositories.pdf

8   APPENDIX

8.1 Interview Questionnaire

Data acquisition, use, and management

What kinds of geospatial data do you use in your organization?

How is that data used; what business processes do they support?

Describe how your organization generally acquires geospatial data?

Are there any proprietary or confidentiality restrictions associated with the use of geospatial data?

Explain in brief how you provide access to your geospatial data to internal and external users.

How do you protect your geospatial data from unauthorized users?

If data is suddenly classified as sensitive or confidential and can no longer be shared, what do you do with archived versions of these data sets?

Are access restrictions on archived data modified?

Your current records management infrastructure

Describe your organizational structure for managing geospatial data and records management.

Can you describe your role?

Describe the geospatial records management system or data repository system that you employ.

Metadata

Are you familiar with state, national, or international standards for geospatial metadata?

Which descriptive metadata elements do you currently use for administration and preservation of geospatial data?

How are the descriptive metadata obtained and entered into the geospatial records management system or data repository?

Are there other descriptive metadata elements that you believe would be useful in your work?

Preserving access

What are your plans for preserving the ability to access geospatial data that you have already acquired or expect to acquire in the future?

How do you preserve capabilities for accessing and managing your current geospatial data given that personnel may turn over and that hardware, operating systems, and software changes over time?

Barriers and enablers

What do you feel are the greatest constraints to ensuring continued access to, and usability of, the geospatial data resources that you have developed to date?

What do you feel are the greatest enablers to ensuring continued access to, and usability of, the geospatial data resources that you have developed to date?

What failures and successes can you share regarding the management and archiving of geospatial data and records?

New tools and resources

What kind of online information or resources about geospatial data management would your organization find useful?

Other Issues

What other issues should be addressed for managing and preserving geospatial data?