



The Earth Institute at Columbia University

Mobilizing the Sciences and Public Policy to Build a Prosperous and Sustainable Future



World Data Center for Human
Interactions in the Environment

Needs Assessment for Managing and Preserving Geospatial Electronic Records: Preliminary Results

Robert R. Downs

CIRESIN

Columbia University

May 13, 2004

The study is being conducted to assess current and future needs for managing Geospatial Electronic Records (GERs) at state, county, and municipal levels of government.

GERs: Geospatial data and related electronic records

Conducted as part of the project, Managing and Preserving Geospatial Electronic Records, funded by the National Historical Publications and Records Commission (NHPRC) of the National Archives and Records Administration (NARA).

Qualitative Research Design

- Protocol to Protect Human Subjects
- Recruit Participants for Interviews
- Interview Participants by Telephone
- Structured Interview Questionnaire
- Analysis of Interview Field Notes

Study Participants

- Advisory Board Nominated 53, Resulting in 49 Unique Candidates
- Interviewed 27 Participants by Telephone
- Distribution of Participants by Professional Role
 - Geospatial Coordinator or Manager: 9
 - Geographical Product Analyst/Developer: 8
 - Geospatial or Planning Dept. Director: 3
 - Archivist: 3
 - Geographical Data Librarian: 2
 - Information Technology/System Director: 2

Distribution of Participants by Organization Type & Government Level

- State Government: 12
- County Government: 4
- Municipal Government: 4
- Educational Institution: 3
- Non-Profit Organization: 3
- Federal Government: 1
- Total 27

General Observations

- Participants represent community of experienced professionals committed to improving geospatial data management practices.
- Reluctance to be interviewed or answer questions reflects concern for compromising security of data, system, or organizational vulnerabilities.

Repeated sentiment about data security: “Since 9/11, all data are confidential”

- Post 9/11, data are screened closely for security and some data are no longer disseminated for security concerns
- Data on locations of infrastructure were available before 9/11, which we cannot distribute anymore
- Prohibited from sharing maps that are public records
 - How can public records be shared if designated confidential?
 - Are some data not confidential?

Interview Questionnaire Categories

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

Observations of Security-Related Requirements for Providing External Access to Geospatial Data

- Written request to order data describing user, data, and intended use
- Formal review of each data access request
- Signed confidentiality agreement to access data
- Protected access and delivery of data
- Contracts with permissions to use data
- Users prohibited from disseminating data

Observations of Security Related Requirements for Providing Internal Access to Geospatial Data

- Internal use by employees only
- Each employee signs confidentiality agreement
- Data cannot be distributed externally to employees
- Control internal distribution to prevent external access

Observed Needs for Sensitive and Confidential Data

- May need to destroy sensitive data
- Data with security restrictions are not disseminated and only accessible by selected individuals
- Each request to access data written specifying user and use
- Report of formal review of request to access data
- Signed confidentiality agreements and/or contracts
- Increases records and resource requirements

Liability and Intellectual Property Requirements Observed

- Identify and record IP rights, citation constraints, and restrictions for access and use when acquiring data
- Obtain access restrictions during acquisition
- Respect license agreements for commercial datasets
- Restrict acquired data for internal use only
- Some organizations have policies to not disseminate data created by others
- Use constraints for each dataset should include statement absolving creator or data provider from liability

Digital Rights Management Requirements Observed

- Provide restricted access to designated layers
- Restrict data to personnel assigned to a specific work group
- Restrict designated data for authorized internal users in organizations with signed Memorandum of Understanding (MOU) for terms of Agreement and Intellectual Property Rights
- Provide secure LAN to prevent unauthorized access from within premises

Examples of Concerns Observed for Confidentiality and Privacy

- “Grappling” with access concerns for data containing confidential information
- Sanitize data or restrict access to entire dataset for any confidentiality concerns
- Remove names and addresses of owners or restrict data
- Locations of protected species, cases where violations have occurred, and orthoimagery

Observations of Requirements for Data Development

- Capabilities to create derivative products from one or more data resources
- Identify aspects of derivative products created from each data source
- Budget to maintain data of continuing value initially acquired for previous programs
- Provide access to most current version of dataset
- Provide discovery metadata in ASCII format with data
- Document techniques developed, assumptions, and conclusions, and for integrating datasets

Needs for Community Resources Observed

- Develop library of case examples, best practices, and recommended technologies for geospatial data storage and delivery
- Document best practices for requesting, logging, classifying, and integrating updates into a distributable database that can be tracked for usage
- Browsable and searchable subject catalogs of free, peer-reviewed, online accessible geospatial data
- Web-browser access to discover data layers, conduct GIS analysis, and mapping
- Peer-review council to review data to be released for public use or included in clearinghouses
- Organizations need to share lessons learned on implementing digital repositories

Archival Needs Observed

- Multiple meanings of “Archive”
- Formal specification for preservation of data
- Education and training on data archiving and preservation
- Designate one work group as archival facility
- Trusted institutional repository where everyone will want to deposit data for preservation and access
- Archive data on CD and DVD media and store one copy off site
- Archive older versions of datasets offline
- Expertise and resources to acquire electronic resources in archive
- Guidelines and standards for archiving electronic records

Observations for Organizational Coordination

- Store Geospatial data in centralized repository
- Establish overall policy to manage geospatial data as records
- Implement a digital repository system to manage data and metadata
- Organization must recognize need to manage geospatial data
- Allocate sufficient staff and cross-train to manage data
- Liaisons from each unit to centralized IT and archive groups
- Revise and document older datasets to be consistent with adopted standards

Observations for Quality Assurance and Documentation

- Track data for current or past projects
- Integrate geospatial data with records management program
- Thematically named folders and files or controlled naming scheme
- Establish minimum documentation requirements and templates
- Policy specifying frequency for onsite and offsite storage of backups and data
- Layer by layer committee review for dissemination and archiving
- Identify common minimum set of metadata elements to be stored in XML-based relational database

Observations of Requirements for Partnerships and Inter-Organizational Cooperation

- Maintain stock written agreements for data sharing
- Develop data sharing agreements between entities specifying sensitivity and security concerns
- For acquisition contracts, require ASCII-Based (SGML, XML) document in standard format
- Form web services partnerships to provide access to and integrate with data served by others



The Earth Institute at Columbia University

Mobilizing the Sciences and Public Policy to Build a Prosperous and Sustainable Future



*World Data Center for Human
Interactions in the Environment*

Questions or Suggestions