



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*

State and Local Government Challenges for Geospatial Data Management and Distribution

Robert R. Downs & Robert S. Chen

Center for International Earth Science Information Network

Columbia University

www.ciesin.columbia.edu

Presentation to

IASSIST 2006, Data in a World of Networked Knowledge

University of Michigan - Ann Arbor, Michigan

May 26, 2006

Abstract

As part of a project investigating requirements for managing and preserving geospatial data and related electronic records, interviews were conducted of thirty-one professionals responsible for managing geospatial data for their organizations. The interviews revealed a range of concerns regarding the management and distribution of geospatial data. Key issues include establishing and maintaining formal agreements, managing intellectual property rights and restrictions associated with the data, protecting sensitive information and the confidentiality of locations revealed by the data, and shielding the organization from potential liabilities resulting from data distribution and use. Many organizations have found innovative ways to address specific issues, but none of those surveyed have fully addressed all of these challenges. Issues identified by the interviews have contributed to the development of a guide for practitioners and a data model identifying information elements to be recorded and maintained when managing geospatial data and related electronic records.

State and Local Government Challenges for Geospatial Data Management and Distribution

- Study Conducted as Part of Larger Project
- Project Overview
- Study Purpose - Needs Assessment
- Methodology to Assess Needs for Managing GERs
- Needs for Geospatial Data Management and Distribution
- Project Products and Resources

Project Overview:

Managing and Preserving Geospatial Electronic Records

- Investigated requirements for state, county, and municipal agencies to manage and preserve geospatial data and related electronic records, especially those generated by GIS software.
- Funded by the National Historical Publications and Records Commission (NHPRC) of the National Archives and Records Administration (NARA)

Project Goals

- Identify and recommend practical and appropriate policies, techniques, standards, and practices to manage geospatial electronic records (GERs)
 - to support their long-term retention and dissemination
 - to facilitate their usability and utility as important information resources of significant historical interest

Project Advisory Board

- **Cheryl Benjamin**, Standards/Data Coordination NY State Cyber Security & Critical Infrastructure Coordination
- **Kevin Glick**, Electronic Records Archivist, Yale Univ. Sterling Memorial Library
- **Bill Guthe**, Educational Tech Center GIS Analyst, Princeton University
- **Geoffrey Huth**, Mgr, Records Services Development NY State Education Dept, Archives
- **Sally Johnson**, Policy Analyst for Statewide Planning Program, State of Rhode Island
- **Patrick McGlamery**, Library Liaison to Geography, University of Connecticut
- **Jennifer O'Neill**, E-Government & GIS Specialist, NY State Education Dept, Archives
- **Theresa Pardo**, Deputy Director, Center for Technology in Government, University of Albany
- **Robert Sandev**, GIS Officer, Division for Ocean Affairs and the Law of the Sea, Office of Legal Affairs, United Nations
- **Doug Schuetz**, GIS Director, Rockland County Planning
- **Bridget Sisk**, Chief, Archives & Records Mgt Section, United Nations Secretariat
- **Terri Spies**, GIS Specialist, NYC Dept of Environmental Protection
- **Larry Thornton**, Bureau Chief, Geographic Information and Analysis, NJ Dept. of Environmental Protection
- **Sam Wear**, GIS Director, Westchester County

Geospatial Data Management Needs Assessment

- Investigated requirements for managing and preserving geospatial data and related electronic records
- Qualitative methodology studied communities of practice
 - Interviews of 31 geospatial data management professionals
- Diverse perspectives, experiences, and practices within community
- Challenges for managing geospatial data and related electronic records

Qualitative Study - Participant Interviews *

- **Project Advisory Board nominated interview candidates**
 - **Each candidate contacted by telephone to request an interview**
 - **Each candidate who agreed scheduled an interview date and time**
 - **Each participant interviewed by telephone**
 - **Interviews lasted up to 1 hour each (usually 30 to 60 minutes)**
 - **Interview guide used for each interview**
 - **Informed consent script**
 - **Semi-structured interview questionnaire**
- * **Human subjects protection protocol approved by Columbia University Institutional Review Board (IRB)**

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?

Demographics of 31 Participants Interviewed

- Gender:
 - Male 26 (84%)
 - Female: 5 (16%)
- Representation by State
 - New York 19 (61%)
 - New Jersey 6 (19%)
 - Connecticut 2 (6%)
 - Rhode Island 2 (6%)
 - Other states 2 (6%)
- Organizational Types:
 - State Government 13 (42%)
 - County Government 5 (16%)
 - Municipal Government 4 (13%)
 - Education Inst 4 (13%)
 - Non-Profit Org 4 (13%)
 - Fed Government 1 (3%)

Professional Categories of Participants Interviewed

- Fields Represented:
 - Planners, Geographers, Systems, Librarians, Archivists
- Diversity of Responsibilities:
 - Directors, Managers, Analysts, Developers
- Position Categories:

– Geospatial Coordinator/Manager	10	(32%)
– Geographical Analyst/ Product Developer	8	(26%)
– Archivist	4	(13%)
– Geographical Data Librarian	3	(10%)
– IS/IT Director	3	(10%)
– Geospatial/Planning Director	3	(10%)

Typical Eclectic Operations

Project-based data acquisition and development of remote sensing, GIS, and other data used for analysis, development of data products, and dissemination

Common Geospatial Data Project Characteristics

- Short deadlines
- Limited resources available
- Competing projects

Unsystematic Treatment of Geospatial Data

- Geospatial data and products used for communication, analysis, and decision-making
- Geospatial data not considered electronic records
- Geospatial data managed with other digital work files
- Unspecified roles of Archivists or Records managers for geospatial data and related electronic records
- Archivists and Records Managers have traditionally managed paper records, not electronic records

Observations of Common Types of Data Loss

- Loss of Non-Archived Data
- Loss of Historical Versions of Data
- Loss of Restricted Data
- Loss of Data Provenance
- Loss of Media Read Capabilities
- Loss of Data Format Read Capabilities

Scenario: Loss of Non-Archived Data

- Symptoms
 - Data were published previously
 - Data downloaded and used for project
 - Data have not been archived
 - Data are no longer available
- Recommended Practice
 - Archive data used for or obtained from projects

Scenario: Loss of Historical Versions of Data

- Symptoms
 - Initial versions of data were published previously
 - New versions of data were produced with updates
 - New versions of data superseded previous versions
 - Previous versions of data are no longer available
- Recommended Practice
 - Record Version along with Title of dataset
 - Identify media or data location with Title and Version

Scenario: Loss of Data Provenance

- Symptoms
 - Data were produced and used
 - Data were not documented adequately
 - Missing information about provider or creator
 - Missing information about creation process
 - Missing information about variables and values
 - Unknown provenance of data prevented reuse
 - Data with unknown origins unavailable for use
- Recommended Practice
 - Archive data creation documentation with data

Scenario: Loss of Media Read Capabilities

- Symptoms
 - Data created and stored on available media
 - Backup tapes
 - Floppy Disks
 - Storage media can no longer be read
 - Media deteriorated
 - Media destroyed
 - Media hardware obsolete
 - Media and hardware available but not documented
- Recommended Practice
 - Inventory media and refresh periodically

Scenario: Loss of Data Read Capabilities

- Symptoms
 - Data created with available software application
 - Data stored in format used with software application
 - Unable to use software that reads data
 - Software not maintained,
 - Software not owned,
 - Software not available
 - Software not documented
 - Software not supported by vendor or produce
- Recommended Practice
 - Convert data to supported formats

Repeated Lessons-Learned

- Assign unique identifier to each dataset to be archived
- Identify relationships to previous versions and predecessors
- Persistent access needed to older data and data products, not just the latest version
- Inconsistencies in practices between agencies
- Identify data containing potential confidentiality issues

Practices Used to Improve Metadata Management With Limited Resources

- Identify minimum metadata required
- Keep required metadata data minimal
- Complete or obtain documentation
- Use database or record-keeping system

Resources Needed for Data Managers

- Database of recommended technologies
- Library of case examples of geospatial data storage/delivery best practices
- Best practices for processing and documenting datasets for distribution
- Advice of experts to help organize data and metadata
- Knowledge of staffing and training requirements for data management

Project Products and Resources

- Geospatial Electronic Records Website
- Guide to Managing Geospatial Electronic Records
- Data Model for Managing and Preserving Geospatial Electronic Records
- Workshop Materials
- Paper: Organizational Needs for Managing Geospatial Data and Related Electronic Records
- Classified References to Standards, Guidelines, and Resources

Key Topics Identified for Guide to Managing Geospatial Electronic Records

- Geospatial Electronic Records Management Policies
- Staffing, Systems, Facilities, and Instruments
- Selection and Appraisal Criteria
- Metadata for Discovery, Admin, & Preservation
- Intellectual Property Rights and Restrictions
- Security, Confidentiality, Privacy, & Ethical Issues
- Disaster Planning and Recovery

Observed Needs for Data Model to Manage Geospatial Electronic Records

- Elements to Meet Standards & Regulations
- Entity-Relationship (ER) Model
- Clustered Design for Full or Partial Use
- Flexible for Implementation in Either:
 - COTS Database Management System Package
 - Distributed Relational Database Management System

Input for Data Model Development

- Elements from Relevant Standards
 - Discovery Metadata
 - Geospatial Metadata
 - Digital Preservation
 - Electronic Records Management
- Elements from Needs Assessment Study
 - Currently Used Elements
 - Elements Needed to Improve Practices

Key Standards Reviewed for Data Model

- **Content Standard for Digital Geospatial Metadata (CSDGM)**, Federal Geographic Data Committee (FGDC). FGDC-STD-001-1998
<http://www.fgdc.gov/metadata/contstan.html>
- **Geographic Information Metadata (ISO 19115)**
- **Reference Model for an Open Archival Information System (OAIS)** Adopted as ISO 14721:2003 <http://www.ccsds.org/documents/650x0b1.pdf>
- **Producer-Archive Interface Methodology Abstract CCSDS 651.0-R-1**
<http://ssdoo.gsfc.nasa.gov/nost/isoas/CCSDS-651.0-R-1-draft.pdf>
- **NARA Records Management. Expanding Acceptable Transfer Requirements: Transfer Instructions for Permanent Electronic Records, Digital Geospatial Data Records.**
- **NARA Code of Federal Regulations Part 1234 Electronic Records Management (5/16/01).**
- **DOD 5015.2-STD (June 19, 2002) C2.T3 Record Metadata & C4.T1 Classified Record Components.** <http://jitc.fhu.disa.mil/recmgt/standards.htm>
- **Information and Documentation – The Dublin Core Metadata Element Set (ISO15836)** <http://www.niso.org/international/SC4/n515.pdf>

Data Model For Managing GERs: Clusters & Tables

- Provenance and Attributes
 - Provenance, Person, Institution, Document, Reference, Property...
- Organization
 - Collection, Series, Aggregate, Series Activity, Retention
- Distribution
 - Dissemination, Distributor, Catalog, Catalog Entry, ...
- Administration
 - Right, Right Held, Right Owner, Access Restriction, Decision, ...
- Physical Properties
 - Media, Location, Event, Environment, Data File, File Part,...

Section from Provenance Table in Database Dictionary

Provenance describes the history and changes that occur during the entire lifecycle of an object and includes information to facilitate discovery within the system.

ProvenanceID	Database generated, Primary Key	Integer not null
AggregateID	Foreign Key, Aggregate Table	Integer not null
Origin	Description of the creation process of the object	String
Title	Formal Title that identifies or describes the object	String
Version	Identification to differentiate object from other versions.	String
PreIngest	Description of pre-ingest processing and custody of object.	String
Creation Date	Date object was completed by Author.	String

Standards and Schemas Matched to the Data Model in Crosswalks

- Cabinet Office, Office of the Envoy. **e-Government Metadata Standard Version 3.0**, April 29, 2004. Available Online: <http://www.govtalk.gov.uk/schemasstandards/metadata.asp>
- Cornwell Management Consultants plc. **Model Requirements for the Management of Electronic Records: MoReq**. IDA Programme of the European Commission. March 2001. Available Online: <http://www.cornwell.co.uk/moreq.html>
- Department of Defense Records Management Program. **DOD 5015.2-STD Design Criteria Standard for Electronic Records Management Software Applications**. Washington, DC. June 19, 2002. Available Online: <http://jitec.fhu.disa.mil/recmgt/standards.html>
- Dublin Core Metadata Initiative. **Dublin Core Metadata Initiative Metadata Terms**. Adopted as Information and documentation – The Dublin Core metadata element set (ISO 15836:2003) and as The Dublin Core Metadata Element Set (ANSI/NISO Z39.85-2001). 2003. Available Online: <http://www.dublincore.org/>
- National Library of Australia. **Preservation Metadata for Digital Collections: Exposure Draft**. October 15, 1999. Available Online: <http://www.nla.gov.au/preserve/pmeta.html>
- National Library of New Zealand. **Metadata Standards Framework, Preservation Metadata**. June 2003. Available Online: http://www.natlib.govt.nz/files/4initiatives_metaschema_revised.pdf
- Online Computer Library Center (OCLC) and Research Libraries Group (RLG). **Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group**. May 2005. <http://www.oclc.org/research/projects/pmwg/premis-final.pdf>



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*

Website on Managing and Preserving Geospatial Electronic Records

www.ciesin.columbia.edu/ger