



Improving Management of Geospatial Electronic Records

Robert S. Chen, Robert R. Downs & Mark G. Becker

Center for International Earth Science Information Network

Columbia University

www.ciesin.columbia.edu

Presentation to the

Association of American Geographers 2005 Annual Meeting

Denver, Colorado

April 9, 2005

Improving Management of GERs

- Project Overview
- “Extending the Boundaries” Workshop
- Study of Needs for Managing GERs
- Needs for Guidance on Practical Approaches
- Needs for Data Model to Manage & Preserve GERs

The project

“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 included identifying and recommending practical and appropriate policies, techniques, standards, and practices

- to manage geospatial electronic records (GERs)
- to support their long-term retention and dissemination and
- to facilitate their usability and utility as important information resources of significant historical interest

Major Tasks of Project

- Establish an Interdisciplinary Advisory Board from Region
- Assess Needs for Managing Geospatial Data and Records
- Review Current Research and Relevant Standards
- Design Data Model to Manage and Preserve Geospatial Data
- Develop Resource Guide and Instructional Materials
- Conduct a Workshop for Agency Practitioners and Managers
- Disseminate Guidance Materials on Web Portal
- Publication of a Research Paper Summarizing Study

CIRESIN Project Team

- **Robert S. Chen**, Deputy Director of CIRESIN and Senior Research Scientist at Columbia University
- **Robert R. Downs**, Senior Digital Archivist and Columbia University Officer of Research
- **Mark Becker**, Geospatial Section Manager and Senior Staff Associate
- **Branko Djapic**, Database Management Systems Analyst
- **Annie Gerard**, CIRESIN Web Designer
- **John Mickelson**, GIS Specialist and Staff Associate
- **John Scialdone**, Senior Metadata Specialist



Project Advisory Board and Workshop Program Committee

- **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

“Extending the Boundaries: A Workshop on Managing and Preserving Geospatial Electronic Records”

(May 13, 2004, Palisades, NY)

Convened to engage practitioners and managers from the geospatial and data archiving communities in collaborative discussions addressing challenges faced by state and local government in managing and preserving geospatial data and electronic records.

Participants

- 25 Professionals from NY, NJ, CT, and RI
- International, Federal, State, County and Local Agencies and Educational Institutions Represented
- Diversified Expertise
 - Geospatial Project Management and Coordination
 - Geographic Product Analysis and Development
 - Planning and Policy
 - Archives
 - Geographical Data Coordination and Libraries

Workshop Recommendations: Policy and Implementation

- Develop criteria for identifying archival records in GIS
- Create a good standard and emergency data/record distribution policy
- Identify secure and confidential data/record based on Freedom of Information
- Create data sharing agreements with all parties (sooner, not later)
- Protection of sensitive/confidential records by destruction is not an option; Cannot destroy records as a way to protect them
- Create secure and confidential data distribution mechanism with secure access controls
- Continuously monitor changes to security/privacy of records throughout the lifecycle of the records (Communication between records creators and preservationists)

Workshop Recommendations: Climate for Change

- New state law to fund and define mission for archives and records management of GIS – for those states without.
- Think outside the GIS/Archiving box (Legal, HR, IT, Project Management, PR, Legislative, Users)
- Integrate activities into existing initiatives so they do not seem to be new work
- Create an enforcement mechanism for each distribution policy
- Involve archivists/records professional in the creation of secure/confidential data distribution mechanisms that produce preservable records (including access logs)
- Create an organization with joint mission: grab turf

Workshop Recommendations: Outreach Coordination

- Continue the work of this MAPGER project
- Web portals for outreach, education, and curricula
- Website of case studies on how to integrate GIS and archives
- Grant program to develop a GIS/archival consortium

Workshop Recommendations: Justification Examples

- Share case studies with cost savings and public benefit
- Centralizing and pooling interagency cooperation vertically
- Fundable project plans
- Put money and resources into distribution
- Manager allocate resources to preserve GIS records

Workshop Recommendations: Communication Among Archives and GIS

- Joint workshops among professional groups
- Provide training and certification program for metadata and archives
- Cross-training of archivists and the GIS field is necessary: archivists to preserve GIS, GIS to produce preservable records, create archivist-type position within GIS organization
- GIS and archivists participate in committee to select GIS records for preservation
- Training on standards for archiving GIS records and techniques for using technology

Workshop Recommendations: Technical Standards and Interoperability

- Promulgate open GIS standards and internet map coordination
- Open source development of standards for archival and GIS records community
- Develop an open non-proprietary GIS format

Needs Assessment Study

- Qualitative Research Study of Communities of Practice
- Interviewed 31 Professionals in Northeast that Use GERs
- Diverse Perspectives and Experiences of Participants
- Identified Challenges for Managing GERs

Participants

- Representation: Primarily Professionals from Northeast States
 - New York, New Jersey, Connecticut, Rhode Island, etc.
- Fields: Planners, Geographers, Systems, Librarians, Archivists
- Responsibilities: Directors, Managers, Analysts, Developers
- Organizational Types: Government, Education, Non-Profit
- Government Levels: Federal, State, County, Local

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

- Data were published previously
- Data have not been archived
- Data are no longer available

Scenario: Loss of Historical Versions of Data

- 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

Scenario: Loss of Data Provenance

- 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
- Data with unknown origins unavailable for use

Scenario: Loss of Media Read Capabilities

- 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

Scenario: Loss of Data Read Capabilities

- 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

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,...
- Organization
 - Collection, Series, Aggregate, Series Activity, Retention, ...
- Distribution
 - Dissemination, Distributor, Catalog, Catalog Entry, ...
- Administration
 - Right Held, Right Owner, Access Restriction, Decision, ...
- Physical Properties
 - Media, Location, Event, Environment, Data File, ...



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