

CODATA Data Archiving Working Group Report

Steve Rossouw, Bill Anderson Co-chairs CODATA Data Archiving Working Group CODATA 2002 30 September 2002, Montréal, Canada



"Digital resources will not survive or remain accessible by accident."

Bernard Smith, European Commission
ICSTI/ICSU/CODATA Digital Preservation Workshop
15 February 2002, Paris, France



Working Group Activities

- Formed at CODATA 2000 in Baveno, Italy
- Preliminary web site established
- Workshop in Pretoria, S. Africa, May 2002
- Annotated list of primary references
- Preliminary classification of issues
- Task Group proposal focusing on developing countries, and
- Collaboration with ICSTI on internet portal

Summary of annotated bibliography (revised 23 Sept 2002)

- Primary and secondary references include:
 - Workshop reports
 - Journals, edited volumes, and standard textbooks
 - Published guidelines and handbooks
 - Government circulars and reports
 - Web sites and portals
 - Standards

Primary references in categories

- Data management
- Data policy
- Scientific Data
- Technical issues
- Technical web resources
- Digital preservation (focus on digital libraries, collections, etc.)

Data preservation and access issues

- Four categories of issues
 - Science
 - Management
 - Policy
 - Technical



- Discipline specific needs and practices of communities;
- Interdisciplinary and pan-disciplinary values, methods



Data Archiving: Scientific Issues

- What are scientific data?
 - OAIS model has reference definitions
- Mandates of different archives differ
- Data quality control and assurance
- Selection and appraisal criteria
 - Value and relevance of data archived
- Language differences
 - Not all data published in one language
- Developing and developed country differences

- Nomenclature / taxonomy
 - Differs inside and across communities
 - Names and concepts change over time (need to save historical contexts)
- Barriers to preservation
 - original data in some fields on paper only
 - original data buried in spreadsheets, databases, documents
- Interdisciplinary work can yield pan-disciplinary, unmanaged data



Practices and procedures of individuals, archival institutions, and communities

Data Archiving: Management Issues

• What is archiving?

4 /

- Relation to other data management functions?
- OAIS model distinguishes issues by
 - archive administration
 - external and community management
- Advocacy needed to secure funding
 - Data management is not science
- Business and organizational models
 - economic and cost, public and private
 - incentives and dis-incentives for populating and maintaining deposits
- Selection and appraisal criteria and prioritization
- Ownership and control

- Planning and requirements issues
 - practices are changing
 - local practices differ
 - mandates and objectives differ
 - what is effective access?
- Applications
 - diversity of customers: scientists, politicians, citizens
- Some operational considerations
 - size
 - diversity: source, formats, documentation
 - time horizon for access
 - changes in data definitions, formats
 - hardware and software obsolescence



Data Archiving: Policy Issues

 Rules, regulations, laws, external to the archive that inform, constrain, and assist management



Data Archiving: Policy Issues

- National, regional & global perspectives
- Cultural ownership of data & preference for use
- Human data privacy & confidentiality
- Environmental data privacy & security
- Intellectual property: protection, limits & exceptions
- Public vs. private data
- Incentives and dis-incentives for managing archive deposits
- National security
- Institutional roles and policies
- Enabling legislation & controlling authorities

- Freedom of information
 - policies, regulations & practices
 - access authorization
- Financing and cost recovery policies
 - economies of scale
 - unfunded mandates
- Rationale for data archiving
 - pure research needs
 - cultural, economic & political needs
- Policy enforcement mechanisms
- Data rights
 - redistribution
 - transformation
 - derivative product rights



- Standards, hardware and software that support data preservation, archiving, and access functions
- Mostly discipline independent

Data Archiving: Technical Issues

- Scientific data and databases are different from literature
 - size and volume differences
 - human readability vs. application access
- Diversity of data types and formats, and media types, formats and standards
- Nomenclature and taxonomy
 - issues apply to the technology itself
- Search capabilities

4 /

- Who need what and to what ends?
- Metadata: difference between access and preservation (OAIS)

- Preservation issues
 - Rapid evolution of technology
 - Information buried in software is hard to maintain and access
 - Information in proprietary formats and commercial databases
- Directories
 - potential user authentication and authorization mechanism
 - potential archive and content discovery mechanism
- Standards: OAIS, Open GIS
 - continuing work is needed
- Interoperability among archives



Task Group Proposal

- "Preservation and Archiving of Scientific and Technical Data in Developing Countries"
 - Improve understanding of S&T data management conditions in developing countries
 - Advance development and adoption of good archiving practices, policies, and tools
 - Provide interdisciplinary forums
 - Build a comprehensive directory of managers, experts, and archives



- Develop and maintain an Internet portal about archiving S&T data and information
 - STI archiving procedures, technologies, standards, and policies
 - Archiving projects and activities
 - Experts points of contact in all countries



Summary

CODATA data archiving activities will pursue opportunities to

- Promote and advance management of S&T data
- Leverage common properties of digital data
- Learn from previous and ongoing experiences with managing growing collections of digital data



Discussion