Digital Archiving: Is the Infrastructure Being Developed?

Gail Hodge
Information International Associates, Inc.

Society for Scholarly Publishing
Baltimore, MD
June 2, 2000

gailhodge@aol.com
Presentation Outline

- Project Background
- Components of Archiving Infrastructure
- Status of Infrastructure Components
- What is Needed?
Project Background

Funding from International Council for Scientific and Technical Information

- 1999 - Study on State of the Practice (co-sponsored by CENDI)
- 1999 - Presentation for World Science Congress
- 2000 - Workshop to bring stakeholders together
- 2000 - Review of the Open Archival Information System Reference Model by ICSTI members
- Raise awareness through publications and presentations
The Digital Information Culture

- More fragmented - ability of producers to “publish” their own information
- More information; harder to tell what should be saved
- Takes the information outside the life cycle that has historically supported archiving
- Impact of multi-media
- Digital information gives impression of permanence
Components of an Archiving Infrastructure

- Culture of Stewardship
- Procedures, Best Practices, Standards and Guidelines
- Technology
Number of different stakeholder groups means multiple cultures

- Authors
- Organizations (provost, corporate information officer, CIO, etc.)
- Technologists (webmasters, ISPs, etc.)
- Publishers (traditional and non-traditional)
- Third party vendors
- National libraries and archives
- Information policy makers
Culture(s), contd.

- Also differences across disciplines and countries
- History of stewardship in the print/physical environment which has not been redefined for the digital environment
- Roles of the stakeholders not yet defined
Major work by national libraries, third parties, data centers and some publishers

- National Library of Canada
- National Library of Australia
- Univ. of Helsinki
- RLG
- OCLC
- NEDLIB
- CEDARS
- Arts and Humanities Data Service
- HighWire Press (LOCKSS)
- Publishers such as Elsevier and American Institute of Physics
- National Library of Medicine
Technologies

- Incorporation of metadata
- Integration in the authoring tools and other aspects of the information life cycle
- Migration instead of emulation now means that simpler is better
What Is Needed?

- Researchers and authors need to take additional responsibility
- Need to support them with tools
- Link to life cycle management
- Procedures need to evolve over time
- Need for common frame of reference
Open Archival Information System Reference Model

- Open Archival Information System RM
  - Developed by the Spatial Data Community
  - ISO Draft due to be released soon
  - Seeks to develop a general reference model for digital (and physical) archiving
  - Common terminology, functions and roles
  - Being used as model by several groups (OCLC, CEDARS, NEDLIB)
  - Gives the various stakeholders a common ground from which to start
What Should Scholarly Publishers Do?

- Become aware of issues and projects
- Get involved in pilot solutions (providing testbeds if possible)
- Plan archiving into new systems (procedures and technologies)
- Emphasize archiving in licenses and agreements
Is the Infrastructure Being Developed?

- Yes
- Moving toward collaboration and common framework
- Options are becoming available
- What remains?
  - Incorporation of principles into tools
  - Research into non-text objects particularly multimedia
  - Raising awareness among authors
Relevant Documents

- “Digital Archiving: The State of the Art, the State of the Practice” April, 1999
  [http://www.icstisti.org/icstisti/whats_new.html]

  [http://www.dlib.org/dlib/january00/01hodge.html]

  [http://www.icstisti.org/icstisti/forum/fo0003.html]