Who Will Keep My Digital Stuff?

Or

Why Even a Non-Extreme Society Needs DSpace

Ann J. Wolpert

Director of Libraries

Massachusetts Institute of Technology

What service attributes would you expect of a repository service?

- Easy to use – submission and retrieval
- Many digital formats
- Cumulative and perpetual
- Describe important details
- Searchable and deliverable
- Persistent identifiers

What organizational attributes would you expect of the service?

- Trusted institution
- Professionally managed
- Distribution capabilities
- Preservation services
- Affordable

Scholarly communication is moving rapidly toward 2030.

- Publication is only one part of the network-enabled “system”
- Disciplines are experimenting
- Traditional outlets are constrained
- New formats present preservation challenges
- Required responsibilities not yet defined

invent@MIT
The HP-MIT Alliance

MITLibraries

Increasing amounts of intellectual output have no print analog.

- The digital genie is out of the bottle in all disciplines.
- Digital works need new, interoperable management and access models.
- Educational content is increasingly digital in format.
- Digital is still frighteningly fragile

The DSpace Repository

- Institutional Repository for MIT faculty's digital research and educational materials
- MIT Libraries - Hewlett Packard Research Labs collaborative development project
- Open Source system
- Federated system
- Preservation archive

invent@MIT
The HP-MIT Alliance

MITLibraries

Institutional Repositories offer part of the solution.

- A tool for faculty and institutions
- Institution-based IP counterweight
- Scholarly and educational material in digital formats
- Cumulative and perpetual
- Open and interoperable

invent@MIT
The HP-MIT Alliance

MITLibraries

Why would MIT and HP invest in building a digital repository?

- Digital preservation
- Asset management in a complex environment
- Partners with different domain expertise
- Mutual interest in layered information management
- Open source philosophy

invent@MIT
The HP-MIT Alliance

MITLibraries

Early adopters have given the system a useful set of examples.

- Sloan School of Management
- Dept. of Ocean Engineering
- Center for Technology, Policy and Industrial Development (CTPID)
- Lab for Information and Decision Systems (LIDS)
- MIT Press –Classic Books

DSpace functions are nominally simple.

- Captures
  - Digital research material in various formats
  - Directly from creators (e.g. faculty)
- Describes
  - Descriptive, technical, rights metadata
- Distributes
  - Via WWW, with necessary access control
- Preserves

Possible content runs the gamut.

- Preprints, articles
- Technical Reports
- Working Papers
- Conference Papers
- E-theses
- Datasets
  - statistical, geospatial, matlab, etc.
- Images
  - visual, scientific, etc.
- Audio files
- Video files
- Learning Objects
- Reformatted digital library collections

DSpace begins with the basics.

- Known/supported
  - TIFF, SGML/XML, AIFF, PDF
- Known/unsupported
  - Microsoft Word, PowerPoint (common)
  - Lotus 1-2-3, Visicalc, WordPerfect (less common)
- Unknown/unsupported
  - One-of-a-kind software program

DSpace is standards based.

- Modular architecture, well-defined APIs
- 100% open source
  - Programmed in java
  - RDBMS and SQL for metadata
- CNRI “handles” for persistent identifiers
- X.509 certificate-based access control
- OpenURL linking
- OAI-PMH for exposing metadata

invent@MIT
The HP-MIT Alliance

MIT Libraries

The technology stack is designed for flexibility and adaptability.

- Apache, Tomcat, OpenSSL/mod_ssl
- Java 1.3, JSP 1.2, Servlet 2.3
- PostgreSQL 7, JDBC (rdbms)
- CNRI Handle System 5 (persistent ids)
- Lucene 1.2 (index/search)
- Jena (RDF History system)
- JUnit (testing), Log4j (logging)
- HP/UX, Linux, Solaris, etc.

DSpace has federation partners outside MIT.

- With support from the Mellon Foundation:
  - Columbia University
  - Ohio State University
  - University of Rochester
  - University of Toronto
  - University of Washington
- With support from other foundations:
  - Cambridge University
  - Cornell University

invent@MIT
The HP-MIT Alliance

MIT Libraries

What's next for DSpace?

- Digital preservation
  - Digital files (e.g. audio, video, image, text)
  - Web sites (e.g. W3C)
  - Software programs
- Personal Archiving strategies
  - DSpace on your laptop
  - Proactive collaboration with content creators
- Federation and economic sustainability

DSpace raises a new set of questions for society, including:

- Can we build a digital formats registry?
- What happens when we have open source, open standards, open access?
- How can the digital work of individuals, units, and enterprises be optimally managed?
- How do you make it easy to give works away?

www.dspace.org

web.mit.edu/dspace