MacKenzie Smith
Associate Director for Technology
MIT Libraries

Institutional Repositories

MIT Libraries

- Institution-based
- Scholarly material in digital formats
- Cumulative and perpetual
- Open and interoperable

The DSpace Repository

MIT Libraries

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

invent@MIT
The HP-MIT Alliance

DSpace

MIT Libraries

- 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

DSpace Offerings

MIT Libraries

- Large-scale, stable, managed long-term storage
- Support for range of digital formats
- Easy-to-use submission process
- Persistent network identifiers
- Access control
- Search and delivery interface
- Digital preservation services

invent@MIT
The HP-MIT Alliance

Possible Content

MIT Libraries

- 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

Challenges

MIT Libraries

- Faculty Acceptance
  - Valuing and trusting an institutional archive
  - Myriad disciplines with different cultures
  - Copyright/IP policies
- Sustainability
  - institutional, financial
- Digital Preservation

Faculty Acceptance

MIT Libraries

- Variety of content
  - Preprints and publications
  - Digital research material
  - Educational material
- Respect for discipline differences
  - Access control, review process, etc.
- Institutional support
  - Broad advocacy
  - Mission relevance

Business Plan

MIT Libraries

- One year, Mellon funded project
- Developed by business consultants, library Transition Team
- Built cost models for running DSpace
- Developed revenue options
  - Core services (free)
  - Premium services (for-fee)

invent@MIT
The HP-MIT Alliance

Digital Preservation

MIT Libraries

- Philosopy
  - Lots of digital material *is already lost*
  - Most digital material is *at risk*
  - Better to have it, do bit preservation than to lose it completely
  - Need to capture as much information as possible to support functional preservation
  - Cost/benefit tradeoffs

Digital Preservation

MIT Libraries

- MIT's commitment levels
  - 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

Digital Preservation

MIT Libraries

- Supported = migration and/or emulation
  - Migration for texts, images, audio, etc.
  - Emulation for software, multimedia?
- Unsupported
  - Bit preservation at minimum
  - Batch migration where possible
    - Commercial conversion services
- Digital Format Registry

Information Model

MIT Libraries

- Communities
- Collections (in communities)
  - Distinct groupings of like items
- Items (in collections)
  - Logical content objects
  - Receive persistent identifier
- Bitstreams (in items)
  - Individual files
  - Receive preservation treatment

invent@MIT
The HP-MIT Alliance

Information Model

MIT Libraries

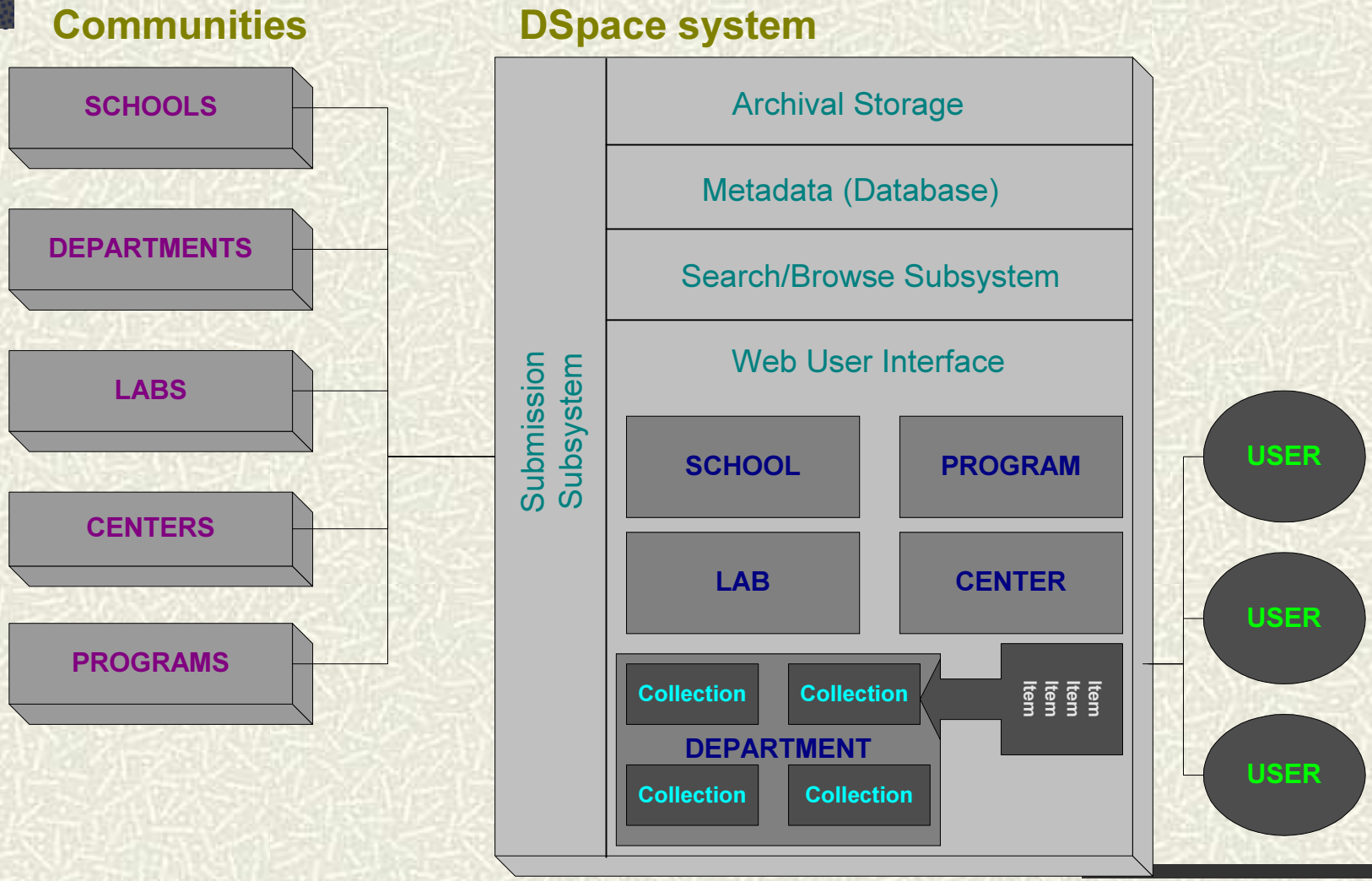
- Versioning
  - Item "versions" can be
    - All instances of a work in different formats
      - E.g. the XML, PDF, and PostScript versions
    - All editions of a work over time
      - Official changes (e.g. addenda or new release)
      - Periodic snapshots (e.g. web sites)
  - Metadata lists all available versions of items

Communities

MIT Libraries

- Departments, Labs, Research Centers, Programs, Schools, etc.
- Localized policy decisions
  - Who can contribute, access material
  - Submission workflow
    - Submitters, approvers, reviewers, editors
  - Collections definition, management
- Communities supply metadata

Communities



MIT Early Adopters

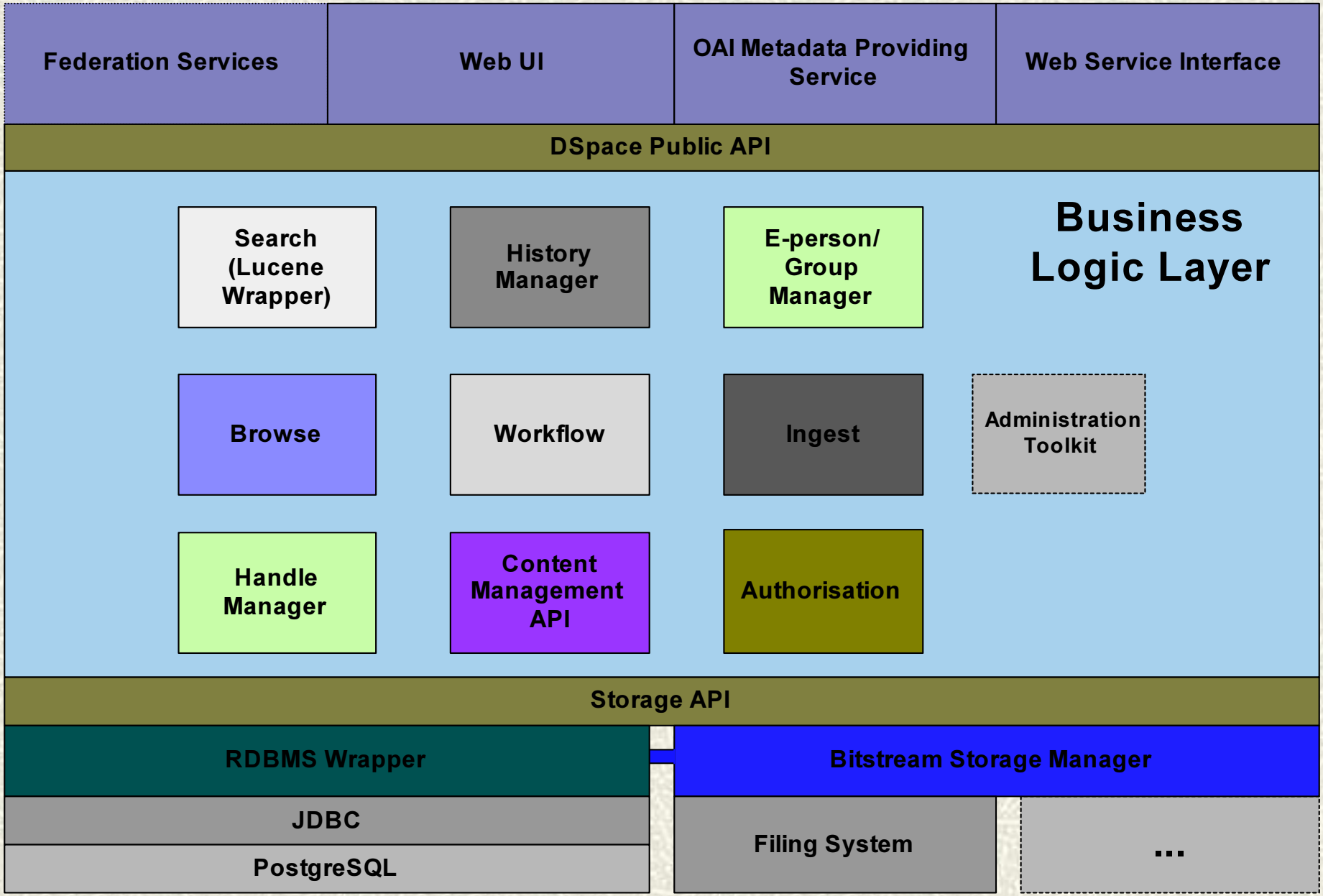
MIT Libraries

- 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 – out-of-print books

Dspace Architecture

MIT Libraries



Standards-based

MIT Libraries

- 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

Technology Stack

MIT Libraries

- 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.

OAIS compliant

- METS AIPs in bitstore
- Designated Community are scholars, researchers
- Knowledge Base
  - Interdisciplinary content
  - Digital archaeology

Metadata

MIT Libraries

- Qualified Dublin Core
  - based on Library Application Profile
- Crosswalk from MARC
  - based on Library of Congress crosswalk
- Minimally effective preservation metadata
- METS-encoded OAIS AIP in bitstore
- Support for collection/community-specific schemas in development (SIMILE)

System Comparison

MIT Libraries

- Extends discipline-based preprint archive model
  - All file formats accepted
  - Preservation commitment
  - Community paradigm
- Differs from Digital Library model
  - e.g. FEDORA, Greenstone, etc.
  - Content is faculty-produced (not library)
  - Responsibility distributed
    - Selection, policies, submission, cataloging, etc.

DSpace Federation

MIT Libraries

- Target audience
  - research libraries, government agencies, cultural heritage institutions (museums, archives)
  - Inside/outside the US
  - Overlapping/complementary research interests

DSpace Federation

MIT Libraries

- Goals
  - Drive DSpace development
    - open source development model
  - Build critical mass of content
    - support useful interoperation
  - Leverage distributed expertise
    - metadata
    - digital preservation

invent@MIT
The HP-MIT Alliance

Federation Benefits

MIT Libraries

- Socio-political
  - Shared direction, leadership, priorities, goals, resources
  - Standards development
    - Putting weight behind “best practices”
    - e.g. W3C, NISO, IETF, ARL/DLF standards
    - Drive commercial developments

Federation Benefits

MIT Libraries

- Technical
  - Virtual collections
    - Networked Digital Library of Theses and Dissertations
    - E.g. Electronic theses
    - Subject-based OAI indexes
  - New publishing models
    - "Overlay" e-journal located at multiple institutions
  - Distributed services
    - Leverage industry services supporting preservation, etc.

Federation Partners

MIT Libraries

- Cambridge University (UK)
- Columbia University (US)
- Cornell University (US)
- Ohio State University (US)
- University of Rochester (US)
- University of Toronto (Canada)
- University of Washington (US)

Schedule

MITLibraries

- MIT public release
  - October 3, 2002
- Open Source to the world (DSpace 1.0)
  - November 4, 2002
- Begin federation
  - Fall 2002

invent@MIT
The HP-MIT Alliance

DSpace: Home - Netscape

File Edit View Go Communicator Help

DSpace

DSpace Home Not log

Search DSpace:

Go

Browse

- Communities/ Collections
- Titles
- Authors
- By Date

Members

- Submit
- Edit Profile
- My DSpace

- About
- Help

DSpace @ MIT Home

What can you find in DSpace?

More Information

- Articles
- Technical Reports
- Working Papers
- Datasets
- Images
- Videos....and more

Search

Enter some text in the box below to search DSpace.

Go

Submit

Submit your digital content to DSpace! To start the submission process, select the button below.

Start Submitting

Authorized Users Only

Browse

You can also browse a DSpace index:

- Communities and collections
- Titles
- Authors
- Dates

My DSpace

You can visit "My DSpace" to resume half-finished submissions and check on the progress of previous submissions.

Visit My DSpace

Authorized Users Only

DSpace Early Ad

Spring-Summer

Communities:

CTPID

Dept. of Ocean Engineering

LIDS

MIT Press

Sloan School

Copyright © 2002 MIT and Hewlett-Packard - Feedback