



as an Institutional Repository

Erja Kajosalo

Librarian for Chemistry & Chemical Engineering

Margret Branschofsky

DSpace User Support Manager



Scholarly Communication

In transition

- **Publications only part of the networked “system”**
- **Traditional outlets constrained (e.g. the “journal crisis)**
- **Disciplines are experimenting**
- **New formats present preservation challenges, some material not “printable”**
- **Responsibilities not yet fully defined**

Institutional Repositories

- **“a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members.”**
- **“commitment to stewardship”**
- **“long-term preservation where appropriate, as well as organization and access or distribution”**
- **“a new channel for structuring the university's contribution to the broader world”**

Clifford Lynch, “Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age”, ARL Bimonthly Report 226, February 2003

DSpace

- **Vision (1999)**

A federated repository that makes available the collective intellectual resources of the world's leading research institutions

- **Mission**

Create a scalable digital archive that preserves and communicates the intellectual output of MIT's faculty and researchers

Support adoption by and federation with other research institutions

What is DSpace?

- **Started as an MIT Libraries - Hewlett Packard Research Labs collaborative development project**
- **Institutional Repository for MIT faculty's digital research and teaching materials**
- **Open Source Software Platform**
- **International Federation of DSpace repositories**

DSpace Federation

- **First international user meeting – March, 2004**
- **Platform for several uses other than IR**
 - **Data sets**
 - **Electronic record management**
 - **Thesis repository**
 - **Publishing platform**
- **Working on plans for governance structure**
- **Plan for encouraging growth of code**

Statistics

- **Federation**
 - **Over 5000 downloads of software (version 1)**
 - **125 universities investigating DSpace for use**
 - **15 universities running production DSpace systems**
- **MIT site:**
 - **Hits per day: 2,563** (without Google robot - 1,763)
 - **Items in DSpace: 3923**
 - **Communities: 11**
 - **Collections: 36**

What DSpace Does?

- **Captures**
 - Digital research materials in various formats
 - Directly from creators (e.g. faculty) or in batches
- **Describes**
 - Descriptive, technical and administrative metadata
- **Distributes**
 - Via WWW, with necessary access controls
- **Preserves**
 - Stable, well-managed, cumulative archive

Possible Content

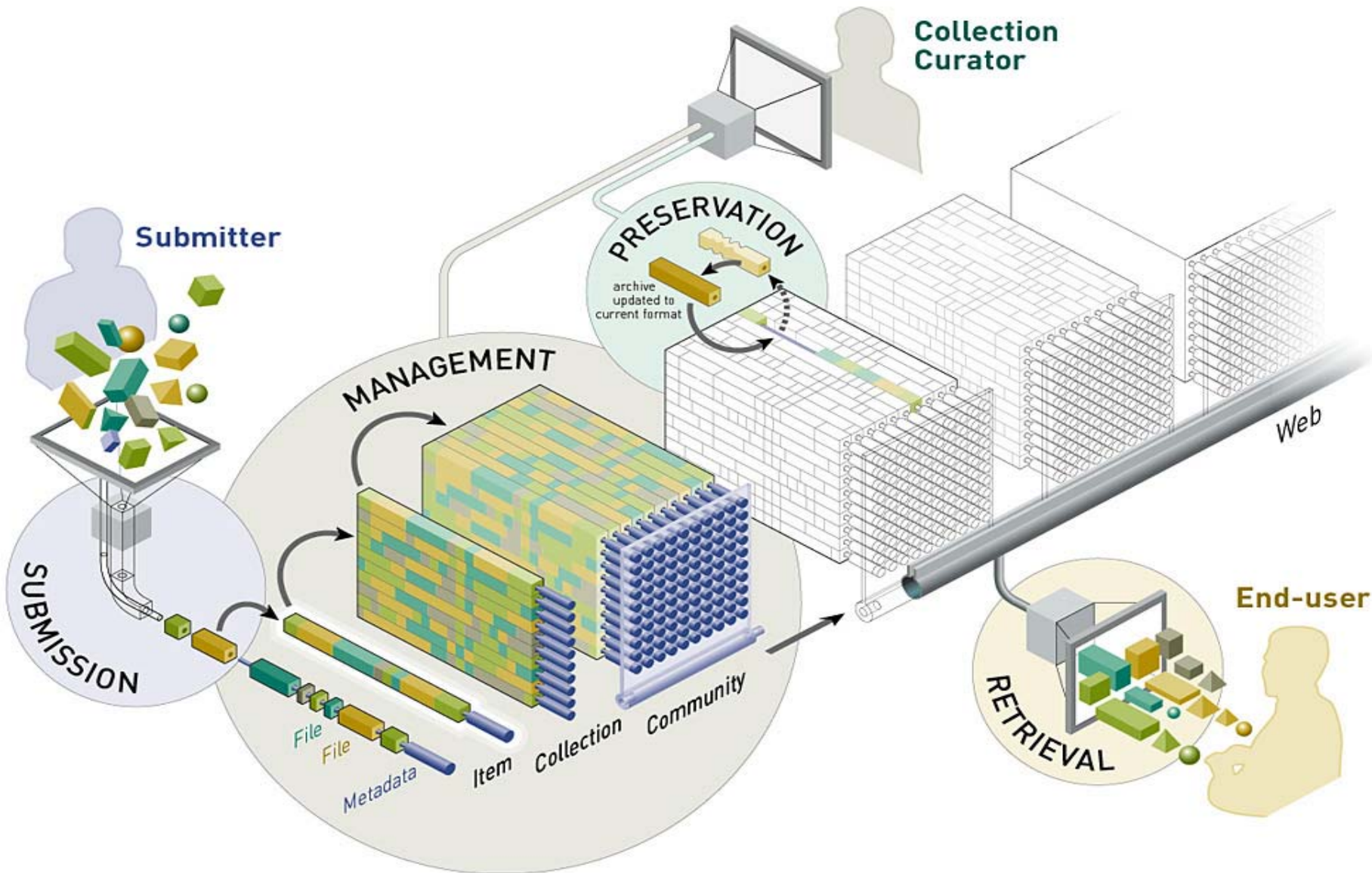
- **Preprints, eprints**
- **Published books, articles**
- **Technical reports**
- **Working papers**
- **Conference papers**
- **E-theses**
- **Images (e.g. Scientific)**
- **Audio/video files**
- **Datasets (statistical, geospatial, scientific)**
- **Databases**
- **Websites**
- **Learning objects**
- **Digitized library collections**

DSpace Information Model

- **Communities**
- **Subcommunities (version 1.2)**
- **Collections**
- **Items (metadata record)**
- **Bitstreams (files)**
 - **Complex objects – multiple files**
 - **Multiple formats - same content**

Metadata

- **Dublin Core with qualifiers**
 - **Based on Library Application Profile**
- **Metadata captures through a form for individual submissions**
- **Metadata unit – cost recovery basis (MIT)**
- **MARC mapped to DC for older collections**
 - **Scanned WPs or TRs or Theses**



Technology Stack

- **Apache, Tomcat, OpenSSL**
- **Java, JSP, Servlets**
- **PostgreSQL, JDBC (rdbms)**
- **CNRI Handle System (persistent ids)**
- **Lucene (index/search)**
- **Jena (RDF History system)**
- **JUnit (testing), Log4j (logging)**
- **HP/UX, Linux, Solaris, etc.**

Standards-based

- **Open Source – openly available to all**
- **CNRI “handles” for persistent identifiers**
- **Open URL linking**
- **OAI-PMH for exposing metadata**
- **SRW for interoperability**
- **METS profile for export**

Digital Preservation

- **Repositories don't “do” preservation**
- **Preservation operations are defined by**
 - **Digital collections in hand**
 - **Cost/benefit tradeoffs**
 - **Local policy**
- **DSpace collects some information that will help in future, e.g., technical metadata**
- **Global Digital Format Registry - <http://hul.harvard.edu/gdfr/>**

Digital Preservation

- **Supported file formats**
 - Open standard, e.g. tiff, txt, pdf, html
 - Confident that preservation techniques possible
 - Batch migration at appropriate time
- **Known (but unsupported) file formats**
 - Proprietary formats, e.g. doc, xls, ppt
 - Batch migration where possible
- **Unsupported file formats**
 - Bit preservation at minimum

Benefits

- **Visibility for institution & individuals**
- **Store variety of digital formats**
- **Rapid distribution of research results**
- **Satisfy funding agencies' requirements to distribute and preserve research results**
- **Long term management & preservation**

Benefits

- **Avoid broken urls when citing work**
 - “40 percent to 50 percent of the URLs referenced in articles in two computing journals were inaccessible within four years”
 - New England Journal of Medicine, Science and Nature: “The prevalence of inactive Internet references grew during those intervals [three months, 15 months and 27 months after publication] from 3.8 percent to 10 percent to 13 percent.”*
- **Persistent network identifiers – national registry service (Handle system - CNRI)**

*Weiss, Rick “On the Web, Research Work Proves Ephemeral: Electronic Archivists Are Playing Catch-Up in Trying to Keep Documents From Landing in History's Dustbin” The Washington Post, November 24, 2003.

Policy Planning

- **Institutional repository – new beast**
- **Policies define and shape repository service**
- **Response to needs of potential users**
- **Needs grow & change over time**

Policy Planning

- **Institutional and library policies**
 - **Intellectual property**
 - **Community-library responsibilities**
 - **Content**
 - **Access**
 - **Metadata**
 - **Preservation**
 - **Withdrawal, retention, versioning**

Lessons Learned

“The policy of being too cautious is the greatest risk of all.”

Jawaharlal Nehru (1889 - 1964)

- **Don't be afraid to make decisions**
- **Don't be afraid to change decisions**
- **Be flexible – interpret broadly**
- **Don't allow policies to impede success**
- **Let users drive service as much as possible**
- **Share policies**
- **Tolerate uncertainty**

FAQ – Intellectual Property

- **Do I retain copyright of my work in DSpace?**
 - Yes, we only require non-exclusive right to distribute and preserve
- **Will I still be able to publish my work after it appears on DSpace?**
 - Yes, for many publishers – can negotiate
- **Can I put previously published work on DSpace?**
 - Yes, if you have retained rights or obtained permission

FAQ - How is this different from my own website?

- **DSpace can complement your website**
 - Point to DSpace from your site
- **DSpace provides certain benefits:**
 - Indexed by Google more frequently than other sites
 - Persistent URLs
 - Search interface – subject, title, author... indexing
 - Additional path for others to locate your work
 - Subscriptions alert users of new work
- **Long term, low-maintenance solution, continuous staffing**

[Advanced Search](#)[Preferences](#)[Language Tools](#)[Search Tips](#)[Web](#)[Images](#)[Groups](#)[Directory](#)[News](#)

Searched the web for **ANAND, Lallit** Results **1 - 10** of about **294**. Search took **0.30** seconds.

[DSpace at MIT: Items for Author **Anand, Lallit**](#)

... Help. About DSpace. DSpace at MIT >. Items for Author **Anand, Lallit** in All of DSpace.
Return to Browse by Author. Sort by Title, Sorting by Date. Showing 1 items. ...

<https://dspace.mit.edu/items-by-author?author=Anand%2C+Lallit> - 9k - [Cached](#) - [Similar pages](#)

[DSpace at MIT: Browse by Author](#)

... Ammar, Mostafa H. (Mostafa Hamed). Amsden, Alice H. **Anand, Lallit**. Ancona, Deborah G. (Deborah Gladstein). Ancona, Deborah G. (Deborah Gladstein). ...

<https://dspace.mit.edu/browse-author?top=Allen%2C+Michael> - 16k - [Cached](#) - [Similar pages](#)

[[More results from https://dspace.mit.edu](https://dspace.mit.edu)]

[SMA - CURRICULUM: People » Faculty](#)

... CS. MIT. Electrical Engineering and Computer Science. **ANAND, Lallit**. AMM&NS.
MIT. Mechanical Engineering. **ANAND, Lallit**. IMST. MIT. Mechanical Engineering. ...

web.mit.edu/sma/curriculum/directories/faculty/faculty-university.html - 101k - [Cached](#) - [Similar pages](#)

[Personal Web Page for **Lallit Anand**](#)

Lallit Anand Professor of Mechanical Engineering address, 1-310. email. **anand@mit.edu**.
phone, (617) 253-1635. ... Biosketch Dr. **Lallit Anand** received his B.Tech. ...

www-me.mit.edu/people/personal/anand.htm - 16k - [Cached](#) - [Similar pages](#)

[Research Topics for **Lallit Anand**](#)

Return to Personal Page, Department of Mechanical Engineering, MIT.

Research topics and projects for **Lallit Anand**. Areas of Interest ...

www-me.mit.edu/people/research/anand.htm - 7k - [Cached](#) - [Similar pages](#)

[[More results from www-me.mit.edu](http://www-me.mit.edu)]

Opportunity for Transformation

New information management paradigm

- **Restore balance**
- **Enable new communication modes**
- **Share in a gift economy**
- **Enhance institution's reputation**
- **Preserve new media**
- **Protect research**
- **Expedite new knowledge transfer**

More Information

- **dspace.mit.edu** (*DSpace at MIT*)
- **dspace.org** (*About DSpace*)
- **dspace-tech@lists.sourceforge.net**
- **dspace-general@mit.edu**