

DSpace at the University of Oregon Library

JQ Johnson, Academic Education Coordinator
University of Oregon Libraries

Online Northwest 2004
20 Feb 2004, 2:45-3:45

Presentation Outline

During 2003, the University of Oregon Library pursued an initiative to investigate and implement a pilot project for an "institutional repository" where members of the UO community could deposit electronic copies of their scholarly works. We evaluated and adopted MIT's DSpace software. This session will discuss our implementation process and will demo the DSpace software we selected.

Agenda:

- Background: What's an "institutional repository"? UO Scholars' Bank
- DSpace, our implementation experience
- Live demo of user interface (network permitting)
- Questions

20 Feb 2004

DSpace at UO

2

Background -- UO



- Medium sized public research university
 - ~1400 faculty; 20,000 students
 - Library: 2.6 million volumes; ~210 FTE (inc. students)
- AAU, ARL, CNI member
- Excellent network infrastructure
- Significant budget problems (3% library budget decrement this year, cutting \$200K serials/year, etc.)

20 Feb 2004

DSpace at UO

3

DSpace adoption timeline

- Spring 2002: participated in CNI and SPARC meetings on Institutional Repositories, strong buy-in at UL level for IR project
- Fall 2002: library IR "initiative" and project team
- April 2003: chose and implemented DSpace as tool for pilot project
- January 2004: assigned project – "Scholars' Bank" – to our newly-refocused "Metadata and Digital Library Services" department
- June 2004 (projected): will move from pilot to full library service

20 Feb 2004

DSpace at UO

4

IR goal setting

- Original goal was to position us to address "crisis in scholarly publishing"
- Shorter term goals identified ranged from institutional self-promotion, service to faculty in making working papers more widely available, opportunity for library subject specialists to connect with research faculty, to internal training opportunity for library staff
- We observed that choice of particular goals heavily influences direction of overall project and choice of technology. "IR" is a Rorschach test.

20 Feb 2004

DSpace at UO

5

Our IR goals for Scholars' Bank



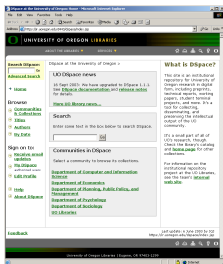
- A place where members of the UO community can deposit their research in digital form
- A tool for collecting, disseminating, and preserving the intellectual output of the UO community
- Metaphor: invest in UO scholarship

20 Feb 2004

DSpace at UO

6

Current state of Scholars' Bank



- Still a pilot project
- Have spent a total of \$150
- Have interested several departments and faculty in publishing, especially
 - PPPM for student terminal projects
 - Economists who already publish in RePEc
 - Several faculty in various disciplines
- But currently only about 100 items, and almost all submissions have been librarian-mediated

20 Feb 2004

DSpace at UO

7

Getting here from there: Some early decisions on IR direction

- Inexpensive; low hanging fruit; "good enough" rather than perfect
 - Low cost hardware and software
 - Expectation that deposit can be unmediated
- UO intellectual output: not just faculty, but not records management
- Self-publication, implying easy user interface
- OAI compliance important to us
- Both short term scholarly access and preservation

20 Feb 2004

DSpace at UO

8

Some typical types of material

Intellectual content

- Working papers
- Preprints
- Postprints/reprints
- Student projects
- Theses & dissertations
- Tech reports
- Supplementary materials (to publications)
- Conference papers
- Computer programs

- Data sets (statistical, GIS)
- Databases
- Grant proposals
- Learning objects
- Archived coursesites and course materials

Material we probably won't collect

- Administrative paperwork
- Committee reports
- Works by non-UO authors (except collective works?)

20 Feb 2004

DSpace at UO

9

Some typical types of material

Document types

- "mostly text", e.g. PDF, Word documents, etc.
- HTML web pages
- Powerpoint files
- Media (video, audio, graphics)
- Archives/packages
- Specialized formats (Mathematica workbook, SPSS system file, etc.)
- and many more...

Formats

- DSpace distinguishes (for preservation):
 - "Supported"
 - "Known"
 - "Unknown"
- "supported" ⇒ forward migration
- We currently are focused mostly on PDF
- DSpace allows multiple bistreams, so multiple versions

20 Feb 2004

DSpace at UO

10

Software we considered (or could have)

- We considered:
 - Virginia Tech ETDdb
 - Southampton E-Prints, <http://software.eprints.org/>
 - MIT's DSpace, <http://www.dspace.org>
- Might also have considered (cf, <http://www.soros.org/openaccess/software/>)
 - Fedora, <http://www.fedora.info/>
 - Bepress (see CDL, <http://repositories.cdlib.org>)
 - Blackboard Content System
 - Other commercial and academic products

20 Feb 2004

DSpace at UO

11

Reasons for picking DSpace

- The "open source" hype. We expected broad based adoption and community development
- Seemed last spring the most powerful available tool
- Rumor was that Eprints sites had found librarian mediation needed for deposit
- We liked the feature set, especially
 - Preservation model (supported, known, ... formats)
 - OAI support
 - Decentralization of control into communities
 - Broader focus than just e-prints
- Implementation seemed likely to be easy given our experience with Linux/apache/mysql

20 Feb 2004

DSpace at UO

12

UO software and hardware in detail

- People: pilot project by ad hoc team; current transition to support by MDLS Department
- Scrounged hardware
 - Hand-me-down server from UO Blackboard system
 - Dell 2400 dual PIII 600MHz/1GB/36GB
 - Network-based tape backup
 - Will add NAS disk space as needed
 - Development/backup server – a random 400MHz/512MB/10GB desktop system

20 Feb 2004

DSpace at UO

13

UO software (cont.)

UO software	MIT recommends	Notes
DSpace 1.1.1		1.2 is coming
Red Hat Linux 9.0 (includes apache 2.0, postgresql 7.3.2)	Any "UNIX-like OS"	We've used RH 8. Will upgrade (RH 3? fedora?)
Sun j2sdk 1.4.1_02	Java 1.3 or 1.4	We'll go to 1.4.2 soon
Jakarta-tomcat 4.1.27	Tomcat 4.0.6 w/apache, otherwise "latest version"	Tomcat 5.0.18?
Postgresql 7.3.2	7.3 or later for unicode	7.4.1?

20 Feb 2004

DSpace at UO

14

UO software (cont.)

- Configuration
 - Standalone tomcat, not mod_webapp or mod_jk2
 - Have CNRI handle; have not yet registered w/OAI
 - Doing SSL in tomcat using a Thawte certificate

20 Feb 2004

DSpace at UO

15

General installation issues

- Moderately complex software system
- Designed for cross-platform portability, but not tested/debugged in enough environments (no resources for substantial testing)
- Too many different pieces, many unbundled. Docs assume familiarity with 3rd party tools.
- Installation documentation is getting better, but needs more non-MIT input and overall bulletproofing
- Upgrade process isn't smooth either

20 Feb 2004

DSpace at UO

16

Initial implementation experience

- Installation went fairly smoothly, but assumed knowledge of Linux, Apache, SSL, Jakarta/tomcat, ant, mod_webapp, PostgreSQL, OAI, CNRI handles, Java Server Pages, Dublin Core, and more
- Consensus on lists at the time was that you needed exactly the version (not necessarily most current) of all these pieces that the documentation recommended. Not following this advice had mixed success for us
- Customizing appearance of our DSpace site was quite easy

20 Feb 2004

DSpace at UO

17

Concrete examples of installation problems

- Required version of various 3rd party tools was unclear. E.g. supposedly Java 1.3 or 1.4 and various versions of Tomcat all worked, but in fact Tomcat 4.1 doesn't run well with Java 1.3, and choice of version interacts with choice of connector and version of mod_webapp
- We decided not to run an apache connector:
 - DSpace docs recommend webapp, but Jakarta project clearly prefers jk2 for tomcat 4.1. Webapp is "deprecated". For tomcat 5, the AJP connector is standard.
 - We couldn't even find a precompiled copy of mod_webapp (the apache end)

20 Feb 2004

DSpace at UO

18

Concrete examples of installation problems (continued)

- Setting up a secure postgresql is a problem; following docs produces a rather insecure setup. See `/var/lib/postgresql/data/pg_hba.conf`
- Documentation does not provide much guidance in setting up SSL with tomcat, particularly with real (non-snakeoil) certs
- Numerous bugs in DSpace 1.0.1. Upgrade to 1.1 reset file revision dates. *Suggestion: wait for v 1.2 rather than upgrading in a month.*

20 Feb 2004

DSpace at UO

19

What we like so far

- System generally will meet our current needs
- Communities provide the right organizational model for us
- Software is simple enough that it was easy to customize appearance, and we can expect to start tuning functionality to our needs
- Basic functionality does meet our needs: OAI, CNRI handles, qDC metadata, file format registry, etc.
- SW is under active development (e.g. Edinburgh, DSpace 1.2)

20 Feb 2004

DSpace at UO

20

What we don't like

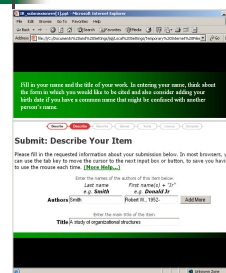
- Inflexible item ingress/metadata collection
 - We want to customize the forms, e.g. add name validation and controlled vocabulary for suggested keywords
 - Authors can't edit own metadata or add bitstreams
 - HTML forms are so retro for file upload (WebDAV?)
- Balance in DSpace between access and archive is a bit too archive-oriented. Examples:
 - HTML documents don't display correctly
 - No provision for non-HTTP delivery e.g. streaming media
 - No easy provision for revisions to working papers
- Communities are more promise than reality, since management can't yet be delegated

20 Feb 2004

DSpace at UO

21

What we're working on fixing



- Develop end user/submitter guides
- Add Radius authentication (either through native Java or webserver delegation), or LDAP authentication?

20 Feb 2004

DSpace at UO

22

What we think DSpace needs



- Plans for 1.2 release (March) seem excellent
- Need additional work on packaging/ installation documentation
- Would be worthwhile to focus for 1.2 on a single platform (e.g. Fedora Core 1, latest Tomcat and Java)
- Need a more active non-MIT development community

20 Feb 2004

DSpace at UO

23

DSpace 1.2

- Beta by March 10? Release by end of March?
- Some proposed new features:
 - Content thumbnail support
 - Full text searching
 - Items shared by multiple collections
 - Delegated administration of communities
 - Improved admin user interface
 - Support for sub-communities (hierarchical)
 - Import/export with METS metadata

20 Feb 2004

DSpace at UO

24

Where to for UO from here?

- Will attend Users Meeting Mar 10&11, and will upgrade to 1.2 soon after release
- Our current issues are less technical than strategic
- We're not yet sure it's easy enough for unmediated submission. Need more testing
- We don't understand long-term preservation. Nobody does
- We're still developing a marketing plan to move from pilot phase to widespread use
- But we think we'll be ready to offer a production service by summer!

20 Feb 2004

DSpace at UO

25

DSpace references

- DSpace Federation, <http://dspace.org>
- DSpace documentation, <http://dspace.org/technology/system-docs/>
- DSpace for Dummies, <http://sunsite.utk.edu/diglib/dspace/>
- DSpace-tech mailing list (essential info!)
- Tomcat/Jakarta documentation, <http://jakarta.apache.org/tomcat/>

20 Feb 2004

DSpace at UO

26

UO References



- UO Library: <http://libweb.uoregon.edu>
- UO Scholars' Bank: <http://scholarsbank.uoregon.edu>
- This presentation: <http://darkwing.uoregon.edu/~jqj/presentations/olnw04/>

For more information: contact
JQ Johnson,
jqj@darkwing.uoregon.edu

20 Feb 2004

DSpace at UO

27