

UNIVERSITY OF EDINBURGH

AND THE

UNIVERSITY OF GLASGOW

THESES ALIVE! / DAEDALUS PROJECTS

**A Report to JISC of the DSpace Users Group Meeting
March 10-11 2004**

Richard Jones and William J Nixon

22 June 2004

Summary

A report of the DSpace Users Group Meeting, March 10-11, 2004.

Correspondence

Correspondence about this report should be sent to Richard Jones (r.d.jones@ed.ac.uk)
or William Nixon (w.nixon@lib.gla.ac.uk)

Document History

Version	Date	Comments
0.9	28 May 2004	Draft for Chris Awre
1.0	June 22 2004	Final draft

Overview

The first DSpace User Group Meeting, held in Boston on the 10-11 March was a useful and very interesting two days.

The first day focussed on DSpace's implementation at various University from the University of Toronto's DSpace to Edinburgh's Research Archive. The various institutions were all at different stages of development. The University of Montreal gave an overview of their French language version of DSpace.

The second day considered broader issues from licensing with the Creative Commons to the Global Digital Format Registry. We then saw an advance preview of DSpace 1.2 and the range of new features which it will include. This was followed by a presentation on the proposed architecture for DSpace 2.x

The remainder of the second day was focussed on Open Source software with a scene setting talk from the Apache Foundation. This led into the broader discussion about the development of DSpace and the Federation which also included contributions from Clifford Lynch.

All of the presentations are available in PowerPoint on the DSpace.org website.

Day One - Presentations

9.00 Welcome – Ann Wolpert, MIT and Nick Wainwright, HP Labs

- Welcome to delegates and introduction of the sponsors: HP and Google
- Nick Wainwright was responsible for the initial proposal to use the BSD licence

9.15 University of Toronto: Federating from within and without: user driven implementation, Gabriela Mircea and Rea Devakos

- Discuss the roles and responsibilities of the involved parties
- Part of CARL (Canadian Association of Research Libraries)
- Part of OCUL (Ontario Council of University Libraries)
- Interested in federated searching
- O-Space: ospace.scholarsportal.info
- T-Space: organisation, technology and resources
- Their collection: 825 items, 32 collections, 409.5Mb
- Why DSpace?
 - Accessibility to materials
 - Digital Preservation
 - Faculty Outreach
 - Sustainable
 - Flexible
- An iterative approach is taken to implement
- Populating the database is mission critical
- Full Text searching is a critical requirement

9.35 Rochester Institute of Technology: Build a digital library and they will come, Marianne Buehler

- 100% availability is the objective
- LDAP implementation
- Their collections are arranged by format

9.55 Kansas University: KUScholarWorks: toward a hybrid organisation for repository items, Holly Mercer and Richard Fyffe, Kansas University

- Working with a small group of early adopters
- ENCompass and metadata harvesting for DSpace
- KUScholarWorks: kuscholarworks.ku.edu
- Adopters include just individual faculty members
- Three different types of community:
 - Formal Communities: mirror the administrative structure
 - Subject Communities
 - Communities of Interest: cross-departmental groupings defined by faculty interest

10.15 University of Montreal: The Erudite implementation of DSpace: translation and organisational issues, Guylaine Beaudry and Luc Grodin, University of Montreal

- Erudite: www.erudit.org, depot.erudite.org
- Four sections:
 - Journals
 - Theses
 - Document and Data Repository
 - Books
- Translation to French (Canadian) of DSpace JSPs and relevant classes
- Displayable Text should be removed from Java classes
- Perhaps implement the JSTL i18n tags
- Creating an inter-institutional repository
 - Dissemination
 - Preservation
 - Batch Upload

10.55 Edinburgh University: Using DSpace for E-Theses at Edinburgh University Library, Richard Jones

- Overview of Theses Alive
- The ETD metadata set developed by RGU, Edinburgh and Glasgow
- The EUL DSpace Theses add-on

10.20 DAEDALUS: DSpace at the University of Glasgow, William Nixon

- The FAIR Programme and JISC
- DAEDALUS Project
- Why both DSpace and EPrints?
- Comparators
- DSpace at Glasgow

11.45 Learning Object Repositories, John Norman, Cambridge University

- Learning Objects and CARET
- Initial resource: tapir.caret.cam.ac.uk
- Using DSpace as a code base
- Collecting materials from multiple repositories and sticking them together for VLEs
- Z39.50 attached to DSpace to drive EndNote
- Custom Teaching application
- Research Group Datasets
- Application interface: IMS DR (OKI DR); SRW (Z39.50)

12.10 Cornell University: Publishing with DSpace, Marcy Rosenkrantz and David Ruddy

- dspace.library.cornell.edu
- Can download the whole catalogue
- DPubS evolved from Dienst (NCSTRL)
- Project Euclid: www.euclid.org
- DPubS:
 - Flexible presentation
 - Simple ingest
 - Flexible subscription Service
 - Modular service architecture (easy to extend)

13.15 Collaboration opportunities between Google and the DSpace Federation, Anurag Acharya

- Researchers do use Google to find scholarly documents
- External users of a system should always be shown some content even if they are not permitted to access all resources
- Issues with crawling DSpace:
 - Finding all the contents
 - What to index: full text, metadata
 - Issues with handles

Track One:

14.00 Track 1: Exploring scalable storage: DSpace and SRB, Chris Fryman

- Proposal to integrate DSpace and SRB
- What if collections could be unlimited in size, stored, replicated and accessible via federated grid technologies
- SRB: Storage Resource Broker
 - Data management infrastructure
 - Utilises data grid & federation technologies
 - Middleware: diverse collection as single logical resource
 - Distributed file system based on client-server architecture
 - Provides ways to access files & computers based on attributes of accessing user
- Data Grid
 - Inter-realm authentication/authorisation
 - Scalability
 - Preservation
 - Client-Server architecture
 - Collective ownership of data
 - Context (metadata) management
 - Abstractions for dealing with heterogeneity
- SRB has multiple API's
 - own security system
 - available for free to academic/research institutions
- SRB Homepage: <http://www.npaci.edu/DICE/SRB/>
- Grid Port Toolkit: <https://gridport.npaci.edu/>
- Data Intensive Computing Environment (DICE): <http://www.npaci.edu/DICE>
- mySRB - Web-based Browser and Query Tool for Storage Resource Broker: <http://www.npaci.edu/dice/srb/mySRB/mySRB.html>
- Presentation: http://libnet.ucsd.edu/dspace/user_group_2004.03.ppt

14.40 Track 1: SRW/U interface to DSpace, Ralph LeVan, OCLC

- SRW: Search & Retrieve Web Service
- WSDL: Web Service Description Language
- SRW provides the option to restrict searches to DSpace communities and collections
- Uses CQL query language
- <http://www.loc.gov/srw>
- <http://www.loc.gov/z3950/srutest.html>
- <http://www.oclc.org/research/software/srw>
- This presentation: <http://staff.oclc.org/~levan/docs/SRWforDSpace.ppt>

15.40 Track 1: DSpace federations and OAI-PMH harvesting: beyond Dublin Core, Henry Jerez, Los Alamos National Library

- MPEG-21 ISO/IEC Committee
- DIDL: Digital Item Declaration Language – Complex object representation
- Files can be harvested from OAI repositories by passing the data in base64 format inside an XML document.
- OAI is the only communications tool used, and the MPEG-21 work occurs in between the client and server ends of the OAI request.

16.20 Track 1: Extending DSpace's repertoire of persistent identifiers, John Kunze, CDL

- Extend DSpace to take advantage of additional persistent identification tools
- The idea of “persistence” is strongly related to the idea of a “service”
- New development: ARK (Archival Resource Key)
- Persistence has to do with the institution, not the technology
- ARKs may be hosted at multiple locations – the hostname is temporary and disposable
- Non semantic naming is superior to semantic naming, since the latter do not age or travel well

Track Two:

14.00 Track 2: The DSpace impact factor: measuring progress in developing a digital institutional repository, Peter Morgan, Cambridge and Julie Walker, MIT

Peter Morgan discussed the importance of evolution and the range of metrics which may be used.

- Why evaluate
 - –to monitor health of system
 - –to demonstrate usage / justify investment
 - –to support case for further investment and development
 - –to establish benchmarks for future assessment
- Overview of DSpace@Cambridge
- Evaluation within the institution
- Evaluation beyond the institution
- To compare progress with other bodies
- To identify community trends
- Range of operational metrics
 - Demographic data
 - How much content is being submitted – is it increasing – at what rate
 - How much content is available
 - How much of it is being used?

Peter concluded with a look at the role of the Federation in Assessment. He also posed a range of questions about what happens next, for instance should the DSpace Federation assume responsibility for standardizing, co-ordinating and disseminating assessment amongst its members? What about co-ordination with non-DSpace repositories such as those using EPrints?

14:40 Track 2: Marketing DSpace and setting policy: MIT's experience, Margret Branschofsky, MIT

Due to time constraints [and interest from the audience] the focus was on marketing rather than policy setting and it was a very useful and helpful presentation which covered:

- Marketing – Pre-production phase
- Phase 1 Early Adopter Program
- Phase 2 Blanket Marketing
- Phase 3 Strategic positioning
- Timing

15:40 Track 2: University of Rochester: enhancing DSpace based on a work-practice study, Nancy Foster and David Lindahl, University of Rochester

This was a fascinating presentation on the use of a work-practice study at the University of Rochester which adapted traditional anthropological methods to:

1. Customize DSpace so that...
 - Faculty want to use it
 - Faculty put a lot of good content into it
2. Provide insights about grey literature
 - How faculty produce and share their own
 - How they find and use others'

16:20 Track 2: Presentation: Teaching digital records archiving with DSpace at the University of Texas, Patricia Galloway

- Austin History Center Projects

Track 3: Google DSpace Federation pilot project logistics

There was also a Track 3 session added to look at Google. Google and Dspace have now launched a joint project which includes Cranfield University from the UK to develop a pilot project for searching University archives.¹

¹ Google and DSpace launch joint project, Open Access Now, <http://www.biomedcentral.com/openaccess/news/?issue=16>

Day Two - 11th March 2004

9.30 Creative Commons, Ben Adida

- Creative Commons
 - Is Not a replacement for copyright
 - Is Not the public domain
 - Is pre-granting permissions
 - Is lowering the cost of collaboration
- 3 Licence types:
 - Human Readable
 - Lawyer Readable
 - Machine Readable
- CC has a plug-in for Mozilla which tells you whether the content is under CC licence (MozCC)
- They have created a tool to include the CC licence selection in the DSpace submission system.
- There is no time-dependent element to the CC licence

10.00 Global Digital Format Registry, Stephen Abrams, Harvard University

- Repository functions are performed on a format-specific basis
- Use Cases:
 - Identification of format
 - Validation of format
 - Transformation of one format to another
 - Characterisation of a format
 - Risk assessment (chance of obsolescence)
 - Delivery/Rendering
- MIME-Type shortfall
 - No syntax and semantic description
 - No requirement for complete disclosure
 - Insufficient granularity
- Unambiguous binding between public identifiers and Representation Information for formats
- Representation Information maps typed formats into meaningful concepts (syntactic/semantic properties)
- <http://hul.harvard.edu/gdfr/>

10.45 SIMILE: Objective, current status and demonstration, Steve Garland, MIT CSAIL

- SIMILE
 - Make the semantic web a reality
 - Provide contextual views of data
 - Allow serendipitous discovery
- Develop flexible semantic web infrastructure (recombinant metadata)
- Joseki (for manipulating RDF)

- <http://simile.mit.edu/>

11.15 DSpace 1.2 demo and DSpace 2.0 Architecture, Robert Tansley

- New Features for 1.2
 - Community hierarchy
 - Thumbnails (auto-generated and stored in a thumbnails bundle)
 - Collection creation wizard
 - Resilient Bitstream ID
 - Delegated administration
 - Items in >1 collection
 - Full Text Indexing
 - Simplified Install
 - METS import/export

At the time of writing [June 2004] DSpace 1.2 BETA 2 is available from SourceForge.

- DSpace 2.x architecture
 - Modularity
 - Digital Preservation
 - Scalability
 - Apache Cocoon for UI Framework
 - Migration from 1.x to 2.x will be non-trivial
 - 2.0 will need to be developed by the community not MIT/HP

A key challenge for the DSpace Federation will be the move to Version 2.x of DSpace.

13.15 The why and the how of Open Source Software: perspectives from the Apache Foundation, Ben Hyde and Stefano Mazzocchi, Apache Foundation

Ben Hyde and Stefano Mazzochi gave a wide ranging and very entertaining insight in the work of the Apache Foundation. They also helped to set the scene on open source development for the DSpace Federation discussion.

- Open up the source code
- OSS backdrop to EPrints and DSpace
- Trusted core developers
- No one is paid, it is all voluntary
- Inverted pyramid model
- Transitioning from commercial space - usually bumpy
- Apache Software Foundation license vs Free Software Licence (GPL)
- "leveraging laziness"

User Group Meeting Summary and Outcomes (DSpace.org)

We discuss what DSpace might be, and how it might develop in the future. Comments come from many areas of the institutions involved: programmers, managers, librarians, academics and more. The key questions for the Federation were: "How to sustain DSpace?" and "What is the best way to get to release 2.0?" At the time of writing a beta release of 1.2 is available on SourceForge. Here we review the points made.

What is DSpace?

Is it

- Just a code base?
- An Institutional Repository?
- The group of institutions using the software?

DSpace is:

- An open source digital asset management system
- Tech platform for institutional repositories
- A federation of digital repositories across multiple academic research institutions
- A production service of MIT libraries for their local research community

Representatives

Various bodies are represented who have looked at DSpace at MIT but other stakeholders have also looked at other ways in which it could be used:

- Research generating organisations (e.g. Universities, research companies)
- For publishing
- Archival organisations
- For preservation

Other uses for DSpace:

- MIT are considering extending the DSpace use to institutional administrative records
- Reading Lists
- Student Coursework (portfolios)
- E-Theses

Concepts for the Future:

- Common Core code with multiple different versions
- A Hybrid model perhaps? Core code with peripheral code

Important Comments

- Tension between sustaining DSpace so that it can serve a range of other needs while the broader Institutional Repositories movement gains pace and momentum
- Clifford Lynch on “The Institutional Repository”: DSpace must survive to bring in the dawn of the Institutional Repository.
- Lynch and Institutional Repositories: IR’s will take place over time. Mackenzie Smith noted that it is a “10 year game not a 2 year one”. Clifford Lynch noted that the primary players in IR’s are the people out talking to and interacting with faculty

and they are quite different from the programmers. Perhaps should disentangle the management of the code base from the paradigm shift of developing IR's.

- Perhaps a distinction between the development of the code base and a role for DSpace to support and encourage IR's and best practice and advocacy?
- Code governance: In open source development you must only govern what you absolutely have to
- Perhaps focus on the role of Institutional Repositories and that could then be spun out
- DSpace must survive to bring in the dawn of the Institutional Repository. It is a "10 year game not a 2 year one".
- Institutional Repositories represent a change in the method of scholarly communication

Questions (and some Answers):

- How much code has been contributed to date?
- To date most of the changes have been made at the edges
- What is an Institutional Repository?
- How do we choose the first trusted core code committers?
- What should DSpace be optimised be to do?
 - Should not be too diluted, especially with the marketplace as it stands now
 - A focus on digital preservation?
 - Keep things simple

Future Developments:

- Virus checking for submission
- More stable DSpace releases more often with shorter paths between the versions
- Communities of applications which focus on particular aspects
- There is a strong desire by the community to more rapidly on to v2.0, but how do we get there. Is there a role for JISC and UK participation?
- Next release around beginning of June [at the time of writing a beta release of 1.2 is available on SourceForge]
- Framework for additional code contributions required

Conclusions

The first DSpace User Group Meeting was a dynamic and interesting one. The opportunity to share experiences with other DSpace implementers. The developments in version 1.2 of DSpace address a range of needs from the user base. The meeting though was very much a clarion cry by the Federation to move to version 2.x and it is now up to the community to work together to realise that vision.

Role for JISC

There may be a role for JISC in the UK to enable and to provide support for the sustained development of DSpace [and Eprints] within the UK HE and FE communities. The decision by some of the Universities within FAIR such as the University's of Glasgow and Edinburgh to work with DSpace to develop their Institutional Repositories in the pursuit of in line with the aims of FAIR.

Further Information

DSpace User Group presentations

<http://www.dspace.org/conference/agenda.html>

DSpace User Group meeting summary

<http://www.dspace.org/conference/conference.html>

Google and DSpace launch joint project

<http://www.biomedcentral.com/openaccess/news/?issue=16>

Last modified on: Tuesday, 22 June 2004