Final Report on the Initial Development of the DSpace Federation
June 30, 2004

EXECUTIVE SUMMARY

In 2002 the MIT Libraries and Hewlett Packard released an open source software system called DSpace™ that enables organizations to collect, manage, preserve, and re-distribute digital content, typically research and teaching material. The system was deployed in late 2002 as a new service of the MIT Libraries, commonly known as an Institutional Repository[1], to the local research community. During 2003-04 a project was undertaken with funding from the Andrew W. Mellon Foundation to promote the adoption of the DSpace platform and the Institutional Repository service model by other institutions, and to in particular to study the ability of six large, technically advanced research universities to adopt the DSpace platform to create their own institutional repositories of digital research material. The institutions that volunteered to participate in this project included Columbia University, Cornell University, Ohio State University, and the Universities of Rochester, Toronto, and Washington. Cambridge University also participated with separate funding[2].

This project, which spanned the first year of DSpace as a free and open source software project, supported a critical stage in the development of both the DSpace platform and the people and institutions that make up the DSpace community. By the end of the project in June 2004, DSpace was running as a live system at approximately fifty organizations world-wide (mainly research universities) and was being implemented, tested, or evaluated by more than one hundred others. The cumulative content of these repositories was approximately sixty thousand items and growing quickly, and significant efforts were underway at a number of participating institutions to develop more effective methods of getting, managing, and preserving this material. A community of open source developers has emerged, and formal structures put in place to enable them to maintain and improve the DSpace platform in the future. Vast improvements have been made to the system, and more are underway or planned. Expertise with the platform has emerged, both technically and among the librarians and other content managers who have curatorial interests in the scholarly material produced by research institutions. Notably, all of the institutions that volunteered to participate in this project have found some use for the platform and are going ahead with an implementation, though to varying degrees and with varying intent.

This is a critical time for the evolution of scholarly communication and for the libraries and archives that manage the scholarly record. Traditional models of communication and

---

[1] There is a lot of literature covering the definition and functionality of Institutional Repositories, and the concept is still evolving. For background, see http://www.arl.org/sparc/IR/ir.html, http://www.dlib.org/dlib/november02/johnson/11johnson.html and http://www.arl.org/newsltr/226/ir.html.
[2] The DSpace@Cambridge project is funded by the Cambridge/MIT Institute to promote the adoption of digital archives and institutional repositories in the UK with Cambridge University serving as an exemplar (http://www.lib.cam.ac.uk/dspace/).

publishing are being challenged, and are beginning to show signs of significant change. Universities, their libraries and archives, are similarly undergoing dramatic changes as the world of information they produce, manage, and consume becomes increasing, predominantly, networked and digital. Platforms like DSpace, while not the only or the final solution to the challenges that these changes present, provide the means for institutions to get involved and explore the policies, legal positions, workflows, business models, preservation practices, and technical requirements for dealing with a digital scholarly record. This project has been extremely successful at getting the DSpace platform to the point where the library and archives community has begun to take ownership of it, and of the myriad opportunities and challenges that it represents.

INTRODUCTION

In November of 2002 the MIT Libraries formally launched DSpace[3], a free, open source software system to capture, manage, preserve, and redistribute digital collections. The venue was a symposium on changes in scholarly communication[4] which reviewed the many challenges facing libraries and archives in supporting digital scholarship and communication, both for current access to research material, and to long-term preservation of that research which forms the scholarly record. MIT had just begun to use the DSpace platform for its own institutional repository[5] service, and was committed to sharing its experiences as broadly as possible, to increase the level of public awareness and debate on these critical issues.

The DSpace project had as its initial vision to create "*a federated repository that makes available the collective intellectual resources of the world's leading research institutions*" and as its mission to both "*create a scalable digital archive that preserves and communicates the intellectual output of MIT's faculty and researcher*" as well as "*support adoption by and federation with other research institutions*".

Because of its mission to promote adoption by institutions other than MIT, the DSpace software was released under a very liberal open source license[6]. The system was made available on the SourceForge website[7] to coincide with the public launch in November 2002, and due to wide publicity it was immediately downloaded by several hundred institutions world-wide. Clearly a collaboration of DSpace implementers was possible, but a concerted effort would be needed to form a coherent Federation of users.

---

[3] http://dspace.org/

[4] http://mitworld.mit.edu/video/72/ and http://mitworld.mit.edu/video/73/

[5] http://www.arl.org/newsltr/226/ir.html

[6] The Berkeley Software Distribution (BSD) license is one that allows for commercialization of the software it covers and so is considered "industry friendly". The license was selected for several reasons: the DSpace system was co-developed with Hewlett-Packard Labs and they might wish to productize a version of it at a later date, and it was an explicit goal of the DSpace project to encourage improvement in commercial products in this area. Both MIT and HP are committed to keeping a version of DSpace in the open source domain, but since it is highly experimental, both technically and socially, it seemed ill-advised to rule out any possible future directions (http://www.opensource.org/licenses/bsd-license.php)

[7] http://sourceforge.net/projects/dspace/

And so, to begin the process of building a community of institutions running institutional repositories using the DSpace platform, the MIT Libraries received a grant from the Andrew W. Mellon Foundation in the fall of 2002, and the DSpace Federation project was launched in January 2003. This project had as its primary goal to establish the DSpace Federation by close collaboration with a small number of universities who would act as testers, advisors, collaborators, and hopefully adopters of the DSpace platform. There were, and are, many other institutions using the DSpace platform, but we felt that having a small number of formal partners would make the task of learning whether and how well DSpace might work outside the MIT environment much more practical and achievable. The initial partners included

- Columbia University
- Cornell University
- Massachusetts Institute of Technology
- Ohio State University
- University of Rochester
- University of Toronto
- University of Washington

This report outlines what the project has accomplished, its major outcomes and discoveries, and a review of what is happening with the DSpace Federation and the Institutional Repository movement in general, one year later.

PROJECT METHODOLOGY AND ACCOMPLISHMENTS

The goal of the project was to test and support DSpace adoption by other institutions, so the work was primarily structured around a series of workshops, site visits, and documentation. Three workshops held at MIT were attended by representatives of the participating institutions, and each institution was visiting by the MIT project advisor to observe progress and provide assistance as appropriate.

The first step in adopting DSpace is to get the technology running locally, so the initial workshop in January 2003 was on technical implementation. This workshop was designed to give participants hands-on experience at installing and customizing the software, as well as a detailed look at the architecture and implementation of the system. HP and MIT developers led the sessions and provided expertise in the DSpace software and related systems needed to create an institutional repository. The workshop aimed to have each institution running the system by the end of the two days, and this goal was achieved. The benefit of having this type of hands-on training and support was valuable enough that the workshop was repeated nine months later in the UK with sponsorship from ERPAnet and the Cambridge/MIT Institute (CMI) for another sixteen UK and European organizations[8]. This second workshop was largely taught by staff from Cambridge University who are working on the CMI-funded DSpace@Cambridge project,

---

[8] http://www.erpanet.org/php/dspace/dspace.htm

also demonstrating the benefit of having a network of collaborators to share this type of work. While installing and customizing DSpace doesn't *require* formal training, as evidenced by the large number of institutions doing without it, clearly the availability of training, technical support, and good documentation can significantly speed up the adoption process, especially for institutions less familiar with this type of technology. One outcome of the project is that we will continue to seek opportunities to partner with organizations like ERPAnet to provide technical training opportunities to the community.

Six months later, in July 2003, MIT hosted the second DSpace workshop on policies and service models. An observation we made from our own experience implementing DSpace at MIT was that the challenges derived from the technology were minor compared to the challenges of building a new production service around digital access and archiving of current research publications and other material. Institutional repositories are unlike familiar print libraries and archives (or their digitized equivalents): they are institutionally-focused and preservation-oriented (like an archive) but provide access to current research and teaching material (like a library). So setting up an institutional repository raises myriad legal, policy, and business questions for which there are few good precedents. To address this, the second workshop included a series of presentations and discussions around the service definition, policy, and business planning aspects of building an institutional repository service, and was attended by several representatives from each of the participating institutions. The format of this workshop also proved quite valuable, and has been replicated at a number of additional workshops on institutional repositories held since then for the library management audience.[9] We anticipate that these issues will continue to require significant and ongoing exploration over the coming years as the role of institutional repositories in academic research organizations becomes clearer.

During the spring and summer of 2003, the MIT project advisor (who is also the MIT project director for DSpace) made a site visit to each of the participating institutions. These visits had the dual purpose of observing first hand the progress of the DSpace project at each institution, and of facilitating that progress by meeting with library staff and administration, making presentations about DSpace to staff and faculty, and consulting on possible solutions to local implementation issues, as appropriate for the particular institution. These visits, in almost every case, helped the staff working with DSpace make significant progress on defining their local service model and increased the momentum of system implementation. It seems to be of great value to institutions grappling with these issues to have examples of policy, service, and business decisions with which to begin their own discussion and debate, even if their eventual decisions are quite different. In other words, it is a barrier to building institutional repository services to start from scratch with a list of issues – having models and case studies is extremely useful.

One early discovery of the project was the need for and benefit of having extensive, accurate documentation for all aspects of DSpace and institutional repository services. In

---

[9] For example, the LEADIRS workshops funded by CMI in the UK, and workshops held at OSU, NERCOMP, CERN, and elsewhere.

order to capture what we knew in a timely way we commissioned a technical writer develop the DSpace Federation website and add significant amounts of material, culled from prior workshops, email archives, internal planning documents, project reports, and contributions from other universities, etc. The existence of this material has enabled many other institutions to get started with DSpace without the need for in-person training or support. The combination of documentation and the various listservs is serving the community's support needs quite well without requiring significant resources from MIT or HP staff.

The final project workshop was originally to have been a gathering of the six institutions for an "all hands" meeting to determine the next steps for the DSpace Federation. At the urging of the participating institutions and by popular demand from other universities the meeting was extended to be the first international DSpace Federation user group meeting, held at MIT in March, 2004[10]. The meeting was open invitation but limited to 100 participants, and was eventually attended by 125 representatives from approximately 50 separate institutions in 9 countries. The two day general meeting was run as a short conference, and included both technical and services tracks. A call for presentation proposals was issued, and twenty-two were accepted. These presentations are now available on the DSpace Federation website (and several under Creative Commons licenses).

The services track was an opportunity for DSpace adopters to share experiences, successes and failures at defining various service models, setting policies, and sustaining their institutional repositories. Included were presentations on using DSpace for open access to journal articles, e-journal publishing, learning object support, electronic theses, electronic records, and more. Many institutions are investing great effort in improving and customizing the DSpace user interface, both for submitting content and for delivering content to the Web in more sophisticated ways, and others are working on support for complex metadata. Several institutions are addressing the problem of marketing the service to faculty, and developing new business models to support it. This is an area where collaboration among institutions and sharing ideas and approaches will be of enormous benefit over the next few years.

The technical track focused on new developments underway: e.g. collaboration with Google for federated searching across the Federation, integration of the DSpace internal file format registry with the planned Global Digital Format Registry[11] in support of digital preservation, support for Creative Commons distribution licenses[12], alternative techniques for assigning persistent identifiers for citations, the SIMILE[13] research project to integrate Semantic Web technology into DSpace for heterogeneous metadata support, and more. Development around the DSpace code base is beginning to become self-sustaining, and projects are starting to find each other for mutually-profitable collaboration around particular problems (e.g. e-thesis support, learning object

---

[10] http://dspace.org/conference/conference.html
[11] http://hul.harvard.edu/gdfr/
[12] http://creativecommons.org/
[13] http://simile.mit.edu/

repositories, e-publishing, and so on). This was a particularly important outcome of the meeting, since it is a necessary prerequisite to achieving a successful open source project.

DSPACE FEDERATION STATUS

One of the most valuable and fascinating presentations of the meeting was by senior members of the Apache Foundation on the governance and mechanics of a highly successful open source software agency. The Apache Foundation oversees some of the world's most successful open source software projects, including the Apache Web server (with nearly 70% of the market share for Web servers) and a set of related technologies in the Web application domain. Their experience in managing open source software, of which DSpace is another example, was both a revelation and an inspiration to the audience. Following the presentation there was a lively discussion of how DSpace might embrace the open source software model, and what the next steps towards building a true open source development community should be for this type of application. It is not certain that systems like DSpace can succeed by following the Apache model exactly (or its close relatives, like the LINUX operating system or the GNU software offerings), but we are fortunate to have their examples to study and apply to our own domain.

The final day of the DSpace Federation user group meeting was closed to the participants of the official Federation project, and reviewed the outcomes of the open meeting to determine how to go forward. The following summary reflects our decisions, and was sent to the DSpace community shortly afterward. It is now available on the project's website:

The DSpace platform is being put to a variety of uses, primarily to create institutional repositories of research publications and other material, but also for other applications (e-thesis repositories, learning object repositories, e-journal publishing, cultural material collections, electronic records management, and so on). Its utility lies in its focus on making content accessible (findable, retrievable) and preserving it over time, and those two things should remain its primary focus going forward, along with developing and perpetuating standards for information and digital object models. There was general validation of a proposed new 2.0 architecture which is more modular and has a more formal API structure that allows other, more context-specific applications to be built on top of DSpace.

The DSpace Federation is now open to all. Anyone can join, everyone can contribute. There are many ways to contribute:

- Become part of the DSpace developer community. We will soon be establishing a group of system architects and developers to share ongoing responsibility for designing, maintaining, and enhancing the DSpace system -- participation is welcome from any institution, and the exact criteria and process for joining the group will be established shortly by a seed group of 5-6 developers from different organizations using DSpace now.

- If you can't contribute programming, there's still lots to do, including: testing and debugging the system (including supplying patches), writing and reviewing documentation, and providing domain expertise to inform functionality that the system should implement (for example, by participating in one of the new SIGs described below). We'll work with the community to establish channels for all these activities within the DSpace Federation framework.

It's time to start thinking about the long-term governance of DSpace outside of MIT or HP (e.g. the social, legal, political, economic, policy, and organizational aspects). We should look at other open source software governance models (e.g. Apache Foundation, Global Grid Forum, W3C, etc.) and develop a plan either to create a new non-profit organization for this purpose, or join an existing one. We should explore various models and policies for transfer of intellectual property - for example, copyright and licensing of code, and the DSpace trademark - to this organization, so that it can be established in such a way that an initial contribution of DSpace intellectual property from MIT and HP is tenable, and so that academic and commercial institutions will be willing to contribute additional IP in the future. As a first step towards this, the current mailing lists and other DSpace Federation services will migrate over to the dspace.org domain from its current home in the mit.edu domain. Work on the DSpace platform and related efforts will continue unabated while this entity is being established under the current, informal governance model.

Institutional Repositories are a high-value, long-term vision, but are still very much works in progress. We need to continue to collaborate on how best to build them, by sharing information about advocacy, marketing, assessment, policies, business plans, and a myriad of other issues that help us understand how institutional repositories will work to best advantage. Since these issues go beyond the DSpace platform to include other organizations, platforms, and views on access and preservation, we need to establish another, more inclusive community with which to collaborate. This community would include DSpace users but not exclusively. SPARC [14] (and possibly other organizations) will be approached to help create this forum, and the DSpace-based institutional repository implementers will help in whatever way we can.

The DSpace Federation will continue to foster new and innovative uses of the DSpace platform by creating a set of Special Interest Groups (SIGs) who can discuss the necessary features and functions of particular applications of the platform. A preliminary list of these might include: institutional repositories, e-thesis repositories, learning object repositories, records managements systems, and publishing systems (open access or other). These will be hosted on the dspace.org website and will be open to all. Institutions using DSpace will ensure that they have cross-representation on these groups and the developers groups so that useful features and functions specified by SIGs are successfully implemented.

---

[14] SPARC is an initiative of the Association of Research Libraries to "create systems that expand information dissemination and use in a networked digital environment while responding to the needs of academe" (http://www.arl.org/sparc)

In the few months since the user group meeting was held, several actions have been taken to follow-up on these decisions. On the technical side,

- A 1.2 version of the DSpace software has been released, which includes code contributions from a number of other institutions, and was entirely tested and debugged by the wider DSpace community.
- An initial group of DSpace "committers" has been formed to take responsibility for the overall system's maintenance and evolution. The group consists of one member each from MIT and HP Labs, and three other individuals from Cambridge University, the University of Glasgow, and OCLC. The small group has developed communication channels and protocols for ongoing decision-making, and is now considering inviting additional members, with a goal of having ten to fifteen members by the end of the year.
- The larger DSpace development community has been provided with a new list (dspace-devel) on which to discuss changes to the platform, both for near-term releases and for a planned 2.0 re-architected and re-factored DSpace system.
- The DSpace lists reflect the level of interest and commitment to embracing DSpace as a true, open source collaboration – the primary technical support list currently has more than 400 subscribers from a wide range of domains, organizations, and countries, and the developer list is currently at 40 active participants. Both of these lists are very active and have contributions primarily *not* from HP and MIT staff (an important factor in measuring how well a community is developing). While these numbers pale in comparison to the Apache software projects, they put DSpace into the realm of open source projects with enough momentum to achieve the goal of self-sustainability.

On the non-technical side of institutional repository development, the idea of creating a community beyond DSpace adopters is receiving a positive response. The DSpace Federation project advisor has worked with SPARC to organize an Institutional Repository User Group meeting to be held in the fall of 2004. The program committee consists of several members of the DSpace Federation (including one from MIT) and representatives from Cal Tech (representing the UK E-Prints platform), the California Digital Library (who use the commercial BePress repository software), and another European member is being sought to represent one of their significant Open Access repository initiative (e.g. the DARE project in the Netherlands[15]). The meeting will be open invitation and a large number of participants are anticipated. There are also conferences planned for the fall on DSpace and/or institutional repositories in general in the UK, India, and Hong Kong. Clearly this is an area on which the research community wants to collaborate further.

DSPACE IN THE EDUCATIONAL TECHNOLOGY DOMAIN

While an important area of policy and advocacy for institutional repositories is providing open access to current research material, the design of the DSpace platform has equal

---

[15] http://www.surf.nl/en/themas/index2.php?oid=7

potential for providing access to, and long-term preservation support for, teaching material. In fact, for many teaching faculty the distinction between research and teaching material is blurred, and the need for a repository to store and manage some types of teaching material and "learning objects" is becoming quite urgent. In order to explore the use of DSpace as a repository for learning objects MIT committed to implementing the Open Knowledge Initiative's [16] Digital Repository OSID (or OKI Service Interface Definition) [17] in DSpace. We assumed that there would be corresponding course management or other educational technology systems that would also implement the Digital Repository OSID so that we could experiment with interoperability between the DSpace platform and those other applications.

As a further example of the general interest in DSpace for teaching material, the University of Arizona has already developed DLearn[18], a production learning object repository that uses DSpace, and is actively collaborating with the DSpace community on the technological changes needed for this type of activity.

The Digital Repository OSID defines an abstract data model and a set of functions which can be performed on the data objects. The DSpace implementation of the OSID is very preliminary, but we are optimistic that more opportunities to refine it will emerge over time. The data model was mapped onto the DSpace digital object model without difficulty, and since the Digital Repository OSID does not specify the encoding or transfer syntax to be used, we chose the METS [19] syntax to encode DSpace digital objects or items. It was further decided to implement the OSID functions using XML-based Web Services[20] since there are a number of educational technology applications at MIT that are not written in the Java programming language and so could not interoperate with Java APIs[21]. The Digital Repository OSID leaves the specific search protocol up to the implementation, so we selected the SRW [22] (Web Service) protocol for remotely searching the DSpace repository.

---

[16] http://web.mit.edu/oki/

[17] http://sourceforge.net/projects/okiproject

[18] http://dlearn.arizona.edu/

[19] METS http://www.loc.gov/standards/mets/ is a proposed standard for the encoding and transmission of digital objects between digital library applications. An alternative choice would have been the IMS Content Package, which is perhaps better known in the educational technology world, but not yet widely supported by actual applications. Yet another option was the MPEG-21 DIDL, and while this has potential in the future for digital object transfer, no one has implemented yet in the educational technology domain. While the choice of METS was a practical one for DSpace and some of the other systems that it currently interoperates with, it is probably not a good long-term choice for interoperating with course management systems. However, while we wait for standards such as the IMS CP to be implemented in that domain, METS is an interim strategy and we hope will be cross-walkable to IMS CP at a later time.

[20] http://www.w3.org/2002/ws/

[21] In particular, MIT has a number of educational technology initiatives funded by the Microsoft iCampus program, all of which use the .NET platform and could only interoperate with a Java-based system like DSpace using standard Web Services.

[22] The Search/Retrieve Web Service, http://www.loc.gov/z3950/agency/zing/srw/, is a successor to the older Z39.50 search and retrieval protocol, a powerful and widely used NISO standard.

The project plan originally called for implementing the OKI Digital Repository OSID during the first half of the project, but during that time the SAKAI[23] project was funded by the Andrew W. Mellon Foundation and the MIT educational technology group made a commitment to migrate MIT's primary course management system (i.e. Stellar) to that platform. The SAKAI re-implementation of Stellar has architectural implications for how DSpace would interoperate with it, and delayed our ability to test our implementation against a real system. Since MIT's SAKAI-based version of Stellar won't be available for another year we designed an implementation of the OKI Digital Repository OSID based on the emerging SAKAI architecture, but do not yet have a system against which to test it[24].

During the project year we found that a number of institutions who have commercial alternatives emerging for the general institutional repository platform (e.g. OSU) are still lacking comparable commercial options for housing their educational content, and so continue to be interested in the DSpace platform as a possible solution to that problem.

While we wait for opportunities to test DSpace interoperability via the OKI OSID, interest in the DSpace platform to support learning object repositories grows rapidly. MIT is investigating using it as such, and has begun a research project funded by the Microsoft iCampus program to archive the course material generated by the OpenCourseWare[25] initiative into the DSpace repository, and to develop the necessary standards and protocols for metadata and content exchange to make that possible. MIT also has further funding to explore the use of DSpace in an educational technology context from the Cambridge/MIT Institute under the DSpace@Cambridge project.

We anticipate that this will be a major area of research and development for the DSpace platform in the next few years, both at MIT and other institutions working in this area.

CONCLUSIONS AND FINDINGS

As it has since its launch in November of 2002, DSpace adoption continues to grow at a rapid rate. With limited formal means of tracking adoption, we are still able to document that there are *at least* fifty DSpace repositories live on the Web today, and forty-eight registered sites (the union of these two sets represents about seventy five organizations world wide). A brief study of the content represented by those fifty live DSpace sites shows that there are approximately sixty thousand digital items available online with this technology, most of it free to the public, and the number is rising steadily. For a project that is not yet two years old, this is quite remarkable progress and demonstrates both the

---

[23] http://www.sakaiproject.org/

[24] None of MIT's currently available course management systems have implemented the OKI OSIDs, so we must look elsewhere to test this functionality.

[25] OpenCourseWare (http://ocw.mit.edu/) is an initiative funded by the Andrew W. Mellon Foundation to publish all of MIT's course material on the public Web.

clear demand for such technology, and the level of interest from the academy to share its research and teaching material, and to get it under long-term curatorial control.

Over the past year the term "DSpace" has come to mean several things:
- An open source software platform for digital asset management with a focus on long-term preservation
- A technology platform with which to build an Institutional Repository service
- A productio n service of the MIT Libraries to its local research community
- A federation of digital repositories across multiple research organizations

Each of these aspects of DSpace is still evolving rapidly as we gain experience with the technology, open source software development strategies, the service model and its related business model, and how to interoperate among repositories and created added-value services over them to best advantage.

A current picture of the DSpace Federation might look like this

DSpace Community

The diagram indicates that the community of open source software developers working on DSpace includes people from beyond the academic research library/archive domain. For a platform like DSpace to survive and thrive without significant ongoing investment from its originators (i.e. HP Labs and the MIT Libraries), the examples from the software industry show us that one of several models is necessary:

- Platform adopters join forces to create a new fee-based membership consortium and designate one member (typically the originator) to provide ongoing development, maintenance and support on behalf of all the members.
- A non-profit or for-profit company takes on the platform development, maintenance and support, and adopters buy the software from the company.
- Adopters support each other in a pure open source software model where little or no money is exchanged, but adopters contribute other resources instead (e.g. software developers, documentation writers, testers, requirements writers, and so on.

The DSpace Federation project provided an opportunity to discuss these options with the group of participating universities, and with the wider DSpace community at the user group meeting. There was a fair degree of consensus that we should not yet opt for a new membership consortium, nor yet look for a company to sell the platform. The group wants to let the open source software development process take its course and see what happens. There are several underlying reasons for this: the academic research community is becoming increasingly dependent on open source software, and increasingly dissatisfied with commercial software, while not entirely understanding the tradeoffs of the open source software approach. The DSpace platform and its development community provide a way to both use and contribute to open source software development in a way that educates adopters about what is really needed.

An excellent example of this comes from the Australian National University. That university's Centre for Sustainable Digital Collections is supporting its involvement as lead institution in the Australian Partnership for Sustainable Repositories (APSR) consortial project[26]. This is a three-year project funded by the Australian Department of Education, Science and Training is focused on building sustainable repository technologies and practices. One of their initial decisions was to spend considerable resources working towards an understanding of the role of open source software and community-based collaboration in building sustainable repositories, and they are partnering with the DSpace Federation to accomplish this goal.

The DSpace Federation is well-launched. The community is forming, and has made its decision to continue to explore the open source software model for the next few years. Whatever the future of the DSpace software and the institutional repository service concept, the Federation community is achieving significant progress in

- building a body of experience within the academic library community in working with an open source software system that came directly out of the digital library domain
- building expertise in the functional requirements for successful digital repositories, which will create an educated consumer base for future commercial offerings in this area

---

[26] http://sts.anu.edu.au/downloads/APSR.pdf

- forming new partnerships with faculty and researchers in understanding their needs and expectations for centralized management and distribution of their digital research and teaching collections
- understanding the costs and benefits of participating in a gift econo my as represented both by the DSpace system and its intended support of Open Access research material
- having the freedom to experiment with new service and business models, including the issues around Open Access distribution of research articles, because of the low cost of entry to building these services
- demonstrating the value of community collaboration in defining new services for academic research universities and their libraries and archives

A recent public policy report on the Information Commons from the Free Expression Policy Project of the NYU School of Law Brennan Center for Justice summarizes the opportunities and threats created by the migration of information and communication to digital technology and online distribution[27]. With the ongoing consolidation of media companies, particularly in academic publishing, and with the ever-increasing specter of pervasive digital rights management in online content distribution channels, the need for alternative approaches, both to the software and to the content policies, is more critical than ever.

MIT intends to continue its active support of the DSpace experiment. We are already finding great benefit from participating in a global collaboration around digital repositories, and have received considerable value back from the community in return for our own initial investment in the software development and other forms of sharing. As the idea of the Intellectual Commons gains definition and momentum, DSpace and its Federation are committed to remaining as an example of finding alternate ways to accomplish new modes of communication in a networked, digital world.

---

[27] The Information Commons: a Public Policy Report by Nancy Kranich, 2004. The Free Expression Policy Project, Brennan Center for Justice at the NYU School of Law.
http://www.fepproject.org/policyreports/InformationCommons.pdf

APPENDIX A: CASE STUDIES

Columbia University

Contact Person: David Millman, dsm@columbia.edu
University URL: www.columbia.edu
DSpace URL: N/A

DSpace Case Study

Columbia University, founded in 1754, is an Ivy League university with approximately 20,000 students and over 3,000 faculty members.

Columbia's Academic Information Services group contributes to an extensive list of interactive services at the university, including the DSpace institutional repository project and EPIC, the Electronic Publishing Initiative at Columbia.

Joining the Federation

Columbia joined the DSpace Federation with several areas of investigation in mind:

- Digital preservation
- Electronic publishing
- Institutional repositories – specifically, electronic publishing and the organizational, policy and technology issues relating to building a repository at Columbia

From the outset, the team identified technology challenges relating to integrating DSpace software with Columbia's internal authorization system, Shibboleth, as well as metadata information tools the library uses.

Columbia's University Librarian and Vice President for Information Services see the DSpace project as a key priority for the library's digital programs.

Building a Service Model

Initially, Columbia planned to build a DSpace service for faculty research, including preprints, working papers and theses. Their DSpace team surveyed faculty informally to see what material might be available for a pilot project.

After consulting with departments and research centers the team now sees DSpace (or whatever software platform they settle on) being used more intensively as a tool for digital preservation and archiving. They also expect to use their institutional repository to

collect a broader range of the University's intellectual assets than might otherwise be possible.

At this point, the team has not developed content guidelines or policies. The service budget is currently absorbed into the general operating budget, and is promoted informally through reference librarians and a small, core team.

DSpace Technology

Columbia has extensive technology resources and staff. The DSpace team's current technical projects include integrating Columbia's local access management, workflow, and acquisition systems into the DSpace software. To date they have modified DSpace core code in the following ways:

- Customized the user interface
- Worked on porting the system to use IBM DS2 as the database
- Substituted Columbia's own authentication system
- Made some temporary adjustments to the DSpace group management system

Their current DSpace service is a test install, with fewer than ten items loaded so far.

Focus on Technology Standards

Columbia is eager for the DSpace Federation to improve the technology architecture and develop an interoperable standard for the platform, so universities can use and remove individual services more easily, and integrate DSpace into their existing technological infrastructure. For example, they have requested a programming interface to make it easier to replace the user interface, and also the ability to distinguish particular bitstreams without using the web interface. The have also recommended changes to the database interaction to make it more generalizable, to the group management, and to the import/export routines.

If they choose to continue developing on the DSpace platform, they would consider contributing code to the central, open code base.

Current Projects

The digital library group at Columbia is currently working on the following projects:

- Electronic publishing projects
- Course management projects (e.g., course reserves)
- Campus-wide image repository
- Long-term digital archiving plan

Evaluating the Federation Project

Members of the Columbia team found the DSpace Federation Project to be very instrumental in deciding how and whether to build an institutional repository.

Looking ahead to how the Federation can help Columbia plan and build their IR, the team suggests these roles for central, federated DSpace organization.

- Developing and clarifying platform requirements.
- Coordinating code enhancements and release management.
- Sharing best practices for service models and marketing.
- Running user meetings.

Lessons Learned

When asked what advice they would offer to other institutions considering running DSpace, the Columbia team offered the following advice:

- Write your own public interface to the software.
- Externalize your metadata handling.
- Get funding upfront before starting building your repository.

The Columbia DSpace team has not yet settled on a software platform for their institutional repository. They continue to evaluate DSpace but have not decided whether to build their institutional repository on the DSpace platform. They are also considering the BePress platform as an alternative to DSpace, and are investigating using a hybrid system including integrated components from these and other locally-developed systems.

Cornell University

Contact Person: Marcie Rosenkrantz
University URL: www.cornell.edu
DSpace URL: http://dspace.library.cornell.edu/

DSpace Case Study

Cornell is an Ivy League, private university and also a federal land-grant institution, with almost 35,000 students, and over 3,000 faculty in its 14 undergraduate, graduate, and professional schools.

The Cornell University Library is one of the twelve largest academic research libraries in the United States, with over 7 million volumes, 65,000 journals and serials, and almost 8 million microfilms. Cornell University Library has developed expertise in digital technologies and managing electronic library resources.

Cornell's focus in building a DSpace service has been two-fold: to create an institutional repository to house faculty research, and to build a new vehicle for scholarly communications and peer-reviewed scholarly publishing. They have brought to the DSpace Federation enormous enthusiasm and energy for creating a new kind of publishing platform.

Joining the DSpace Federation Project

Library staff at Cornell joined the DSpace Federation Project to learn about the software platform and to investigate its use as a vehicle for open access publishing. DSpace would be part of a wider effort at the university to reduce the library's costs for journals and books.

Cornell has already developed an online institutional repository called DPubS, or Digital Publishing System, which was used to create Project Euclid (http://ProjectEuclid.org) to publish proprietary serial literature in math and statistics. There are plans to create links between DPubS and DSpace, potentially using DSpace for access control or even as the underlying repository for DPubS.

Current plans for DSpace are to include the following types of content:
- Faculty research work
- Graduate, professional degree and undergraduate theses
- Student journals
- Open access publishing using the Internet-First University Press (IFUP) community in DSpace

There are currently about 40 items in the DSpace system at Cornell.

Resources and Staffing

The DSpace project at Cornell is run primarily with existing staff members and promoted by university administrators, most notably, Bob Cooke, Dean of the Faculty. There is one staff member (half-time) who does technical work and system administration. Existing library staff set up DSpace communities, authorize submissions, edit metadata, etc.

The editorial board of the Internet-First University Press (IFUP) provides oversight for the DSpace team's work.

Establishing Policies

There are no set policies for the Cornell DSpace service. The team deliberately avoided implementing specific policies in the hopes of reducing the barriers to participation as much as possible.

Further, there are no content guidelines. Submitters are encouraged to use PDF files, but the service accepts other formats with caveats.

Anyone in the community can submit content to the service – faculty, student assistants, administrators, editors, webmasters, etc. Library staff review metadata for theses but not for other content types.

Budgeting and Service Planning

Cornell's DSpace participation was funded by a private foundation for a two-year initial period. For the next six months, this grant money will support the DSpace service, and after that date, the library systems budget will absorb the costs of running DSpace.

There are currently no plans to add premium, fee-based services to recover operating costs.

Early Adopters

Cornell has five early adopter communities so far:
- Cornell University East Asia Papers
- Cornell University Graduate School (online theses)
- Cornell University Library
- Internet-First University Press (IFUP)
- System Dynamics

The bulk of submissions are in the online theses and the Internet-First University Press communities.

The IFUP offers full texts of new and out-of-print books for free on the Web at http://dspace.library.cornell.edu/handle/1813/62. Readers can buy printed copies through a print-on-demand system.

DSpace Technology

One feature enhancement the Cornell team would like to see is the ability to have different metadata for different communities or types of files – for example, datasets, videos, etc. Their organization does not have the programming resources to add this functionality themselves.

While the team is enthusiastic about open source and would like someday to contribute code to DSpace, this is also their first experience being part of an open source community and their technology resources are tight.

Evaluating the DSpace Federation Project

The Cornell team gave high marks to the DSpace Federation Project experience, particularly the meetings and the chance to share experiences with other DSpace teams. They found it to be enormously helpful in deciding how to build their institutional repository and they look to the Federation and other team members to learn best practices for service models and marketing.

When asked to provide a potential mission statement for the DSpace Federation, the Cornell team suggests that it focus on continuing to build the software platform, "for maintaining, preserving, and disseminating the scholarly output of our faculty, staff, and students."

Lessons Learned

The Cornell team offers valuable advice to other universities just starting to build institutional repositories:
- Don't depend on the "if you build it they will come" model.
- Get people on board first.
- Concentrate on early adopters – demonstrate success there first.
- Define your goals carefully.
- Define success for your organization.

Ohio State University

Contact Person: Laura Tull, tull.9@osu.edu
University URL: www.osu.edu
DSpace URL: http://dspace.lib.ohio-state.edu/index.jsp

DSpace Case Study

Ohio State University has over 49,000 undergraduates and more than 13,000 graduate students. Its main Columbus campus is the second largest campus in the United States.

In 2001, the university's Distance Learning/Continuing Education Committee identified the need for a repository to house faculty research and course materials. Out of this project came the Ohio State University Knowledge Bank, an institutional repository that serves the entire OSU community of students and scholars.

Knowledge Bank is one of the OSU Library's top priorities and will launch in Fall of 2004, and DSpace will be a component of the Knowledge Bank infrastructure for the institutional repository.

Ohio's State and Local Repositories

Ohio State created its Knowledge Bank institutional repository to store, preserve, and share the valuable digital content created by its academic faculty. This campus-wide institutional repository uses the DSpace platform.

Separately, the OhioLINK consortium (http://www.ohiolink.edu/), of which OSU is a member, developed a state-wide institutional repository using Documentum. This repository also contains a Digital Media Center (http://dmc.ohiolink.edu/), among other collections from around the state of Ohio.

OSU can now choose which collections are more appropriate for which repository. For example, OSU houses its electronic theses and dissertations at OhioLINK and has used its Digital Media Center for image collections. In the future, some student works (e.g. undergraduate honors theses) will be in Knowledge Bank, on the DSpace platform.

Identifying Campus Needs and Marketing the Service

When the Libraries' Digital Initiatives Steering Committee set out to create Knowledge Bank, they surveyed all of the digital projects on campus. Using this list, they contacted faculty involved in digital projects to see how DSpace could meet their needs.

The OSU team also held one-on-one informal meetings with key instructional technology personnel on campus to identify needs the online repository would have to meet.
The team promotes the service in presentations at department meetings and through individual contacts with Deans and department heads. Knowledge Bank is also promoted by the CIO and in the Library's strategic plan. (Select presentations are available online at http://www.lib.ohio-state.edu/Kbinfo/#presentations.

One key factor in identifying potential online communities is finding groups on campus who have published content on paper but not yet online. Their files exist in published format, in PDF for example, and can easily be submitted to DSpace.

Finally, the library team learned about DSpace by testing the system with the library's own content. To learn about building a community and seed the service with content, the team created a community for the OSU Libraries itself, to house the OSU library newsletter.

Staffing

The Knowledge Bank team consists of existing library staff plus a few new positions:
- Program Manager, who markets the service to faculty and communities
- Systems Librarian, existing position, providing user support and some marketing, part time (20%) on DSpace
- Metadata Specialist, new position for Knowledge Bank
- Head of Access, Support and Accounting Department, an existing position in the library that provides metadata consultation to new communities
- Systems Developer/Engineer, existing position in campus IT group
- Senior System Developer, new position in the library for Knowledge Bank programming

Each position has additional responsibilities outside the Knowledge Bank service.

Technology

The OSU team started the project with a very simple hardware configuration on a lone PC and has upgraded their hardware and storage plans as the service model takes shape.

IT staff on campus did the initial installation. As the Knowledge Bank grows to include software developers, they will do code customization for the service and build the user interface.

Programming resources have been limited so far, so the team has prioritized feature requests and customizations to the DSpace code, making good use of their scarce resources. For any code they do develop, they plan to share code with the DSpace Federation for others to share the benefits of their work.

One key area for future work is to integrate the repository with the campus authentication system. This work would be specific to OSU and not shared with other federators.

Current Status

The Knowledge Bank DSpace service currently has three online communities up and running:
- Learning Objects Research (6 items)
- Ohio State University Libraries – Friends of the Libraries Newsletter (3 items)

- Olentangy River Wetland Research Park (22 items)

Future online communities include a collection of World War II videos, images and text, which library staff will create with the OSU History Department.

Budgeting

The Libraries received $400,000 in one-time Ohio Board of Regents' Research Challenge funding to support Knowledge Bank through June 2005. The university support is also underwriting two years' funding for the Knowledge Bank Program Manager. The library's operating budge absorbs costs for part-time contributions for other Knowledge Bank employees.

To date, the OSU team has incurred no single, large expense specifically related to DSpace implementation, other than a significant amount of library personnel time spent exploring the institutional repository concept and defining a course of action. The OSU Offices of the CIO have partnered with the Libraries on the Knowledge Bank and contributed a server and IT staff time to support their DSpace pilot projects.

To offset some of the cost of running Knowledge Bank, the Libraries are considering offering digitization and metadata services to help communities to submit content to the Knowledge Bank. They plan to use central funding for startup and testing of such services, which would need to become fee-based for cost recovery once the central funding has been used.

Working with the DSpace Federation

The OSU team suggested a parallel between the Federation's role and that of a Library System Users Group, which suggests feature enhancements and votes on priorities for development.

The DSpace Federation got high marks for providing the foundation OSU's team needed to get their service started more quickly than they might have done on their own.

Among the lessons OSU's team learned – and would like to share with others – is to survey faculty before building the service and also set expectations about what the service can offer. Because faculty want one service to meet all their needs, the DSpace team on campus has to be clear about what DSpace can and cannot do for them (especially when marketing the repository and trying to recruit communities).

University of Rochester

Contact Person: Susan Gibbons, SGibbons@library.rochester.edu
University URL: www.rochester.edu
DSpace URL: https://dspace.lib.rochester.edu/index.jsp

DSpace Case Study

Founded in 1850, the University of Rochester is a private research university with over 8,600 students and 1,140 faculty members. Among its noted divisions are the Eastman School of Music and the School of Medicine and Dentistry. Its Institute of Optics, founded in 1929, is considered one of the nation's premier optics schools and is a leader in basic optical research and theory. Rochester's electronic imaging research and teaching programs are considered to be among the most innovative in the country.

With more than 3.3 million volumes, Rochester's library ranks among the top 50 of academic research libraries in the United States and Canada.

Joining the DSpace Federation

The university Provost invited the Libraries to investigate building an institutional repository and faculty e-archive in 2000. After examining all the available software platforms, the Rochester team selected DSpace before joining the DSpace Federation Project.

DSpace will serve as Rochester's only institution-wide repository platform, but other campus organization will choose various software products to house smaller collections, intranets, and extranets.

Rochester views DSpace as the "showcase" repository for the institution.

DSpace Service Model

Rochester's DSpace service accepts all faculty-sponsored collections, with the following guidelines:
- Work must be produced, submitted, or sponsored by a faculty, research, or staff member.
- Work must be scholarly, educational, or research oriented.
- Work should be ready for public dissemination (not a black archive).

Early collections will include working papers, image collections, out-of-print faculty titles, and conference proceedings.

As of Spring 2004, there were 135 items in Rochester's DSpace service, housed in five early adopter communities:
- Library collections (historical images)

- Computer Science department
- Visual & Cultural Studies
- Electrical Engineering department
- University Publications Office

The Library plans eventually to offer fee-based services such as metadata at the deposit level, large storage requirements, and format migration.

The service has not formally launched yet. Once the user interface changes are complete, the team will do a public launch with public relations and marketing activities on campus and in the Rochester area. (Current marketing materials are available at http://www.library.rochester.edu/index.cfm?PAGE=1346.)

Faculty Work Practice Study

Rochester is currently conducting an IMLS grant to study differences by academic discipline in faculty use of gray literature. This work-practice study is led by an anthropologist and includes interviews and observations.

Part of the grant work includes designing modifications to DSpace to better align it with faculty needs. The findings of this grant (ending in September 2004) will impact the design and service model of Rochester's DSpace service. (See http://www.library.rochester.edu/index.cfm?Page=IMLS.)

Resources and Staffing

There are three staff members on the DSpace team at UR:
- Project Director, who spends 1/3 of her time on DSpace
- DSpace Administrator, a programmer who devotes 100% of his time to DSpace, doing system administration and Java programming
- Bibliographer, who spends 1/3 of her time serving as faculty liaison to DSpace

In addition, the library has a Usability Team, with 7 staff members who conduct usability tests on the DSpace user interface. A Design Team (graphic designer and web master) do design modifications to the UI based on results of usability testing.

A DSpace Advisory Board meets once a month, and is comprised of 6 staff members who make decisions about policies, collections, for-fee services, etc.

The team wishes they had the resources to hire an additional Java programmer because administering the service takes up most of the existing programmer's time.

Initial costs for establishing a DSpace service were covered in part by a $25k gift from the Provost's office and remaining costs were absorbed into the operating budget. The largest costs were equipment costs and hiring a DSpace Administrator/Programmer.

Establishing Policies

Rochester's DSpace Policy Board used the policies MIT and other institutions developed as a starting point for their DSpace policies. Rochester accepts any format, but follows MIT's example of designating preservation levels: supported, unknown, not-supported.

Faculty and staff will submit their own content to DSpace and enter their own metadata. The library staff have created some of the early collections, though, including a collection of historical images.

UR's policies are available online at
http://www.library.rochester.edu/index.cfm?PAGE=1275

DSpace Technology

The technical features Rochester would like to see added to future releases of DSpace include:
- Thumbnail support – Rochester has extensive image repositories that require this feature.
- Collection-level administrators
- Drag-drop uploading
- Better user interface
- Full-text searching
- Access restrictions at an item, rather than collection level

Rochester has a programmer on staff and is willing to contribute code to the shared DSpace code base. This is their first experience participating in an open source development model. While they miss having a customer service staff to call on, they are excited to be able to modify the platform's source code as needed.

Experience of the DSpace Federation Project

The University of Rochester was already determined to establish an institutional repository before it joined the DSpace Federation. So the Federation served Rochester primarily with technical support and contacts with other institutions to share experiences. It also offered good PR for Rochester's institutional repository service, which helped the team to garner additional support from the university.

Rochester's Project Director notes that she wishes there had been more meetings among the federators this year. She never felt she had a good sense what the other federators were doing. She notes that it "Would have been nice to have had more time to get all of our DSpace installations up, with some content and then some OAI services built on top in order to demonstrate all the potential."

The Rochester team was particularly helpful in sharing its experiences and research with other federators.

Role of the DSpace Federation

Rochester looks to the Federation to support, maintain, and coordinate the DSpace code base and to share best practices for service models and marketing among other organizations using DSpace. A third priority would be advocacy of institutional repositories and related issues.

Also, they hope the Federation will take the lead in working on preservation issues. Rather than having each institution tackle preservation issues individually, they would like to see the federated group do it together.

When asked to provide a mission statement for the Federation, Rochester's project director suggests the following: "To ensure the integrity and development of the DSpace code base and to provide DSpace users with the support they need to successfully implement and develop an institutional repository with DSpace. 'Support' would include manuals, workshops, user groups, list serves, and preservation guidelines."

Lessons Learned

Rochester's DSpace team has shared the benefits of their research and experience with other DSpace federators, both informally and at group meetings. They offer the following list of lessons learned to other institutions embarking on building an institutional repository.

- The most complex parts of the project deal not with the technology, but with the politics, people and policies.
- It takes a great deal longer to get an institutional repository up and running and to get content "flowing" into it than most would like to admit.
- Consider this to be a five-year project before it begins to run like a more established service, rather than a new one.

Rochester's DSpace team has strong university support, and a strong staffing model. They are eager to create a successful service and happy to assist other universities embarking on similar projects.

University of Toronto

Contact Person: Rea Devakos
University URL: http://www.utoronto.ca/
DSpace URLs:
- TSpace: http://tspace.utoronto.edu
- OSpace: http://ospace.scholarsportal.info

DSpace Case Study

Founded in 1827, the University of Toronto has approximately 38,000 full-time students and over 10,000 faculty. Its library has over 15 million holdings – making it one of the top 5 research libraries in North America.

The university library has focused great attention on DSpace this year. There are two DSpace projects currently underway at the University of Toronto:
- TSpace, the university's DSpace service
- OSpace, a service for OCUL, the Ontario Consortium of University Libraries, that hosts research material from a variety of Canadian

Launched in 2003, these projects currently have 8 communities, 32 collections and 825 items. The asset store has over 400MB of data. With these DSpace trial services up and running, the U of T is focused on moving beyond the test phase and populating the repositories with content.

Joining the DSpace Federation

The U of T joined the DSpace Federation because its goals were similar to theirs: to learn about institutional repositories and how best to implement them. They were interested in a DSpace model of a self-sustaining and scalable implementation, as well as the self-submission feature for adding content to the repository. They were eager to learn from other federators and to share their experiences as well.

The U of T libraries are committed to using DSpace for now but have not ruled out using other systems as new needs arise.

Service Models

TSpace accepts scholarly research work in variety of content formats. Faculty research projects include: working papers, self-archiving, images, websites, datasets, audio files, and previously unpublished works. The collections focus on largely complete works, but there are faculty requests to use DSpace for works-in-progress where content would be gradually unveiled. There is a pilot project now to add work-in-progress content to TSpace.

There are no current plans to add theses to TSpace. There are, however, discussions about creating a student instance of DSpace at University of Toronto which might include student e-portfolios.

For the OSpace/OCUL project, they are testing the use of DSpace functionality for a province-wide repository of information literacy learning objects.

Staffing

The U of T Library use existing staffing to run DSpace, using a combination of applied costs and contract staff. The team consists of an IT director, a programmer who also sets up new communities and support users, a project leader, and information specialists. As at MIT, the metadata entry and quality control are offered as cost-recovery, fee-based services.

Signing Up Communities

The U of T DSpace team assessed campus needs by doing semi-structured interviews with potential early adopter communities and influential faculty and administrators. The team felt that having a few successful demonstration projects would be critical to deploying this type of IT project on campus.

The TSpace team uses a variety of strategies to find new communities and keep in touch with their early adopters:
- Interviews of existing communities about needs, progress, expectations.
- Presentations to early adopters and existing library committees. These presentations also help to assess faculty needs and gather suggestions.
- Email list for early adopters and those interested in learning about new developments.
- Periodic contact by phone, in person, and over e-mail. Every 3 months, TSpace staff informally contact community users to assess interest level, status, etc.
- When visiting new communities, the team sends a technical staff member with a non-technical staff member, to help answer any technical questions that arise.

Creating Policies

A Policy working group sets TSpace service policies and includes the project programmer (who also provides user support), service coordinator, and head of the Library's IT group. The group reviews the policies of other repositories and similar services, consults with library staff and users, and drafts TSpace policies.

The Reference Services and the Council of Campus Libraries committees review the policies before they go to the Library Council. The Chief Librarian oversees all major policy decisions.

Content Submissions

Content guidelines include the following:
- Work to be included must be scholarly, complete and not ephemeral.

- TSpace accepts all content formats, to accommodate the vast range of faculty output.
- Communities can restrict access to materials to university users only.
- TSpace records are exported to the library's catalog system.

Generally speaking, student assistants do the bulk of content submissions, but faculty and administrative staff do some as well.

Library staff generally do not add or review metadata for content submissions. The TSpace team discusses metadata issues and standards with communities and users.

To see if the best content model is to have communities always submit their own content, the library has experimented with some forms of assistance for loading content. The library team loads content from a community's web site, exporting pages into item submission forms. Library staff then review and edit the data.

Budgeting

The costs of running DSpace are currently absorbed into the library's operating budget. The team has not sought grant monies yet because they haven't found an appropriate grant for their needs.

Marketing

The U of T service has a two-pronged and evolving marketing strategy:

1. **Focused, targeted viral marketing campaign**
   By building a small demonstration repository of about 500 items of representative content, they hope to show a variety of formats, disciplines and policies. They marketed the demo repository using word-of-mouth, existing contacts, "chaining" contacts – getting suggested contacts from others and targeted presentations at campus-wide events or committee meetings.

2. **Campus launch**
   Marketing strategies for launching TSpace include articles in the university newspaper, library newsletter articles, a revamped website, and inclusion in campus planning processes, training sessions and presentations. The team also plans a formal launch campaign in conjunction with other digital library initiatives at U of T.

Community response has been positive so far. These features resonate particularly well with faculty:

- Community and author control of the submission process and access
- Persistent identifiers
- Open access
- Uniform framework for institutional presence
- "Easy Content Management"

29

DSpace Technology

After working with the DSpace technology for over a year, the U of T team identified a few technical issues of concern:
- The submission process, while simple, is somewhat tedious to complete.
- More tools are needed to make it easier to add content to DSpace. (They have offered to work on programming these features and sharing code with others.)

The group has worked with other open source tools but has not contributed code to those tools, which include Apache, Perl, MySQL, and Tomcat. They are, however, eager to contribute code to DSpace.

Next Steps

Some of the challenges they identify are specific to the university and others are shared by all DSpace federators. Primary among them is reaching critical mass – how to populate the service with content and spread awareness in the scholarly community. Sustainability is vital – both in terms of the self-submission process for growing the service and in terms of resources devoted to running an institutional repository. Additional areas of focus include the following:
- Populating the service with content
- Analyzing customers – who are the heaviest users, the most committed, etc.
- Focus on the user interface and design
- Exploring use of the Creative Commons License for DSpace

The TSpace team at U of T has taken an active role in the DSpace Federation, by articulating the challenges they face and sharing their strategies for creating a viable service.

Future of the Federation

The TSpace team describe "Access to MIT and HP resources" one of the highlights of the DSpace Federation Project, along with support from other federators and sharing experiences through email and at user meetings.

They describe these project experiences as most useful to share with other DSpace services: policy decisions, ways to sign up more communities, and strategies for moving beyond the early adopter phase.

In the future, they look to the Federation to offer more work on federated search across DSpace services.

University of Washington

Contact Person: Bill Jordan, bjordan@u.washington.edu
University URL: www.washington.edu
DSpace URL: https://digital.lib.washington.edu/dspace/

DSpace Case Study

The University of Washington (UW) is a large public research university in Seattle serving over 42,000 full-time students. The UW Libraries system includes 5 million volumes and an equal number in microform, as well as online serials and other formats. The Libraries system has a particular focus in online technologies, in concert with the university's support for digital scholarship.

Joining the DSpace Federation

The Libraries' initial motives for joining the DSpace federation project included faculty requests and the Libraries' interest in preserving born-digital material. They also found DSpace's community model and open source distribution to be important for their applications, which would need to interact with other online initiatives on campus, including the DigitalWell media archive project.

Last year, the university conducted a scholars meeting to assess the state of digital research at the university and the online technologies needed to collect, preserve and distribute faculty research.

At UW, the first DSpace community in development is a collection of Buddhist Manuscripts. The Libraries are also looking for grant funding for a project to gather public health gray literature.

In the future, the library staff hopes to use DSpace as a repository layer in e-publishing projects. Their technical staff may need to modify the central code base to uncouple visual rendering from the repository layer.

Resources and Staffing

The DSpace project at UW is staffed by library staff drawn from the Digital Initiatives group, Information Technology Services, subject specialists/liaisons, and the Cataloging department. They may add a DSpace Project Manager as they build the service.

Current collections include about 800 documents, hosted on a hardware configuration that includes 1 development, 1 test, and 10 production servers. The hardware and software are managed by information technology staff who also work on other IT projects at the university.
The DSpace team is debating how and whether to offer additional services such as metadata and schema/thesaurus support, as well as quality control of metadata. Currently

only library staff submit content, so as additional communities join the project, these issues will need to be addressed.

Establishing Policies

Currently a Steering Group of library administrators, technical staff, and department liaisons sets DSpace policies. There is some discussion of adding a faculty advisory group, possibly under the larger umbrella of university-wide support for digital scholarship.

Content guidelines have not been formally established. In the course of developing the first two communities, the group developed a Rights Agreement for the service. The library's Metadata Implementation Group did some consulting for the early communities' submissions, providing tips for creating and reviewing metadata.

Budgeting and Service Planning

The DSpace project's budget has up until now come out of the library's operating budget. The team has sought grant funding for the DSpace service, and the potential new hire of a DSpace Project Manager will be funded through endowment income.

Planning for the service's business model and cost recovery is under discussion now. The DSpace team is following the Cornell Digital Consulting & Production Services model to see if their cost recovery service will prove to be viable.

The team also plans to do some market analysis and planning. They are selecting early adopters for the technology by choosing projects that might teach the DSpace team different aspects of running an institutional repository and supporting digital scholarship.

Several articles on campus have helped to promote DSpace at UW, but they do not constitute a formal promotion/marketing plan:
- http://admin.urel.washington.edu/uweek/archives/issue/uweek_story_small.asp?id=975
- http://www.lib.washington.edu/digitalscholar/other-sites.html

DSpace Technology

The UW team identified several pieces of additional software functionality they would like to have in DSpace:
- A feature where DSpace easily ingests the contents of an existing website
- Rendering, or display, of the repository's contents
- Administrator user interface
- Group management improvements

This is the UW team's first open source project, so they are still a bit unclear about functionality they can contribute. If they do develop these features they are willing to contribute their code to the open source code base.

Evaluating the DSpace Federation Project

The DSpace Federation Project proved valuable to the University of Washington team for several reasons. They had had the idea of building a service to capture and preserve the university's digital assets for several years. DSpace gave them the platform they'd been looking for: "DSpace gave us something concrete to work from, although we have not progressed as far and fast as we would have hoped."

The UW team see these as the primary roles of a DSpace Federation:
- Support, maintenance, and ongoing evolution of the DSpace code base
- Sharing best practices among federation members
- Advocacy of institutional repositories

They also look to the Federation for central code coordination and to run a user group for mailing lists and organization, if not necessarily for large group meetings.