

Building a Business Plan for DSpace, MIT Libraries' Digital Institutional Repository

Mary R. Barton and Julie Harford Walker
MIT Libraries, DSpace Project
77 Massachusetts Avenue, Building 14S-216
Cambridge, Massachusetts 02139-4307
Fax: (617) 253-8894
mbarton@mit.edu
(617) 253-7746
jhwalker@mit.edu
(617) 258-8303

Key features References; **Figures** 1 **Tables** 1 2

Abstract

This paper presents an overview of the methodology and results of the MIT Libraries' business plan development project for DSpace (<http://www.dspace.org/>), MIT's digital institutional repository. The introductory section includes a description of DSpace, the objectives of the business plan project, and the current status of the DSpace project. The methodology section explains the process and tools with which the business plan was developed. The remainder of the paper describes the results of the business plan project, including the DSpace service definition, the cost model, potential funding sources, and future DSpace plans.

Keywords: digital libraries, institutional repository, scholarly communication

1.0 Introduction

DSpace (<http://www.dspace.org/>) is MIT Libraries' innovative institutional digital repository designed to manage, host, preserve, and enable distribution of the scholarly output of MIT's faculty. Developed as a joint research project of the MIT Libraries and Hewlett-Packard (HP) through invent@MIT, the HP-MIT Alliance, it reflects MIT's mission to "generate, disseminate and preserve knowledge" and provides MIT faculty with a stable long-term storage and content management system to house their digitally formatted work.

DSpace was developed in response to expressed faculty needs for an easy-to-use, dependable repository that accommodates a broad range of formats. It is intended to provide a solid foundation for the collection of digital material from around the Institute. MIT Libraries hold a non-exclusive license to distribute and preserve items but do not own the DSpace content.

DSpace has been developed by the MIT Libraries as a continuation of their mission to collect, make available, and preserve important scholarly material of all kinds, especially that of MIT's own faculty and research community. The Libraries are working to extend their services into the digital era, to reflect current trends in scholarly communication and education, and to offer new means of distributing research material that are enabled by network technology. It is incumbent upon libraries to develop strategic and economic plans for the preservation and usability of those resources over time. Analogous to print materials, digital library initiatives require cost-benefit considerations that must be carefully weighed against other library priorities.

An award from the Andrew W. Mellon Foundation in April 2000 provided the MIT Libraries with a unique opportunity to examine business issues as they apply to the Libraries' newly developed DSpace service. In this report we will share an overview of how MIT Libraries developed a business plan to transform its DSpace research project into a sustainable technology platform and service administered by MIT Libraries and adopted by the Institute's producers and consumers of digital scholarly materials. We also will share the Libraries' plan for a collaborative Federation of DSpace systems at other universities and cultural heritage institutions.

This plan is a snapshot of our business strategy in Fall 2002 and one that is evolving as more users adopt the service. It reflects the organizational and technical infrastructure and resources of MIT. Other institutions interested in implementing the DSpace system will be able to scale it to match the needs and resources of their own organizations, creating a repository with the scope of digital formats and the scale of content that address their constituent's needs. MIT Libraries designed DSpace as an Open Source (<http://www.opensource.org/>) system that could be licensed without cost or restriction to other institutions to extend the DSpace technology into "a federation of systems [that] makes available the collective intellectual resources of the world's leading research institutions."

1.1 DSpace Status

Four early adopter research Communities at MIT have been using the DSpace service since February 2002 to test and demonstrate the viability of the submission process. DSpace was officially launched to Communities throughout MIT in September 2002 when worldwide access to the contents also became available. The DSpace source code was made available in November 2002 to Federation partners and other interested research universities under an Open Source license.

2.0 Business Plan Methodology

We developed the DSpace business plan collaboratively with the DSpace Transition Planning Group, a group of leaders from MIT Libraries staff, the DSpace development team and MIT's Information Systems organization. The Transition Planning Group was charged with outlining the operational, policy, marketing, management and staffing plans

for the Libraries' DSpace service. The work of this group is represented heavily throughout our business plan.

We should also note that because DSpace is the first service of its kind, we chose to be conservative in our modeling. Rather than attempt to make speculative projections about the growth of the system, our model allows funding to provide for limited, manageable growth. Our model also reflects the need to manage costs and the impact of DSpace on existing library resources, a key factor in the design of the DSpace service definition. We were unable to find literature on any other business model for a comparable, library based, institutional service.

Another group, the Early Adopter Librarians task force, staffed with Subject Specialists from each of the Libraries, provided additional insight into DSpace's potential user base from their direct interactions with faculty members. This group also was key in understanding the organizational impact that the implementation of DSpace would have on Libraries staff.

Our review of digital library and related solutions, both for-profit and non-profit, revealed a highly fragmented market with individual initiatives selecting different aspects of the problem to address and different means to achieve solutions. We learned that the need for institutional repositories is widespread based on the multitude of new projects and the inquiries we received from other universities. We were able to develop an understanding of how institutionally based digital repositories like DSpace can best respond to the needs of institutional constituents. We confirmed that DSpace is relatively unique in that there are no other institutionally based repositories that invite the breadth of file formats, have digital preservation as an explicit objective, and provide a flexible, decentralized community-based submission process.

We administered a survey to the tenure track faculty of MIT in order to learn about their perceptions and anticipated use of DSpace. Respondents were roughly representative of the overall Institute's department and tenure mix. Our findings served as a means to validate aspects of the DSpace Service Definition, which is covered in more detail later in this report.

We developed a cost model to capture the full economic cost of operating DSpace including staff impact, space, hardware and other Libraries resources, only some of which will result in differential cash flows. The remaining costs are important to capture for MIT Libraries' planning purposes. We gathered data from the Transition Planning Group staffing model, HP, MIT Libraries' records and MIT central accounting. Our model is designed to reflect the costs of operating DSpace; therefore we specifically ignore system development costs. The costs of implementation within MIT Libraries are also ignored. Because DSpace is the first of its kind our costs of implementation would not necessarily be replicated by federating institutions.

We also examined possible funding opportunities to offset the costs of operating the system. We took into consideration MIT Libraries' goals to make the system freely

accessible to both submitters and consumers of content. Using the service definition and cost model as a guide to the potential scope and scale of the system, we sought to maintain a balance between rapidly advancing the system and minimizing costs.

3.0 The DSpace Service

The DSpace Service is divided into two main areas:

- **Core Services**, which are available at no charge to Community members and consumers of DSpace content; and
- **Premium Services**, which are specialized services designed to meet the extraordinary needs of Community members and may be offered on a fee for service basis.

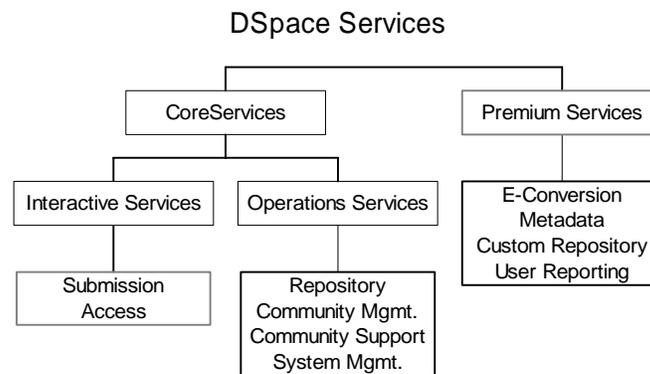


Figure 1. DSpace Services

DSpace Core Services are comprised of two distinct but interconnected service elements, Interactive Services and Operations Services. DSpace Interactive Services offer a fully functional system that allows DSpace Community members and consumers of DSpace content to accomplish all tasks necessary to submit and access items in DSpace, as applicable. Additionally, MIT Libraries provide Operations Services to host and preserve faculty materials, establish and deliver ongoing support for DSpace Communities, respond to customer inquiries, and supply system monitoring, back up, and recovery.

From early feedback that we received, we anticipated that DSpace communities or individual faculty members may put extraordinary demands on the service such as sizeable storage requirements or assistance with specialized metadata creation. MIT Libraries plan to offer **Premium Services** to ensure that DSpace offers a full set of resources to meet faculty and researcher's needs and to manage the impact of these exceptional resource requirements on Libraries staff and DSpace resources. MIT Libraries reserves the right to introduce Premium Services fees as needed to aid in cost recovery. We do not speculate on the growth of the services, they will be introduced as library resources allow. Neither do we speculate on revenues generated, as it would be

an inappropriate representation of these services as for-profit when the fees charges would merely cover the costs of providing them.

Further definition of the Premium Services and market validation of the demand is being explored as DSpace is adopted campus-wide. Libraries traditionally have fostered open accessibility to information resources. Fee-based Premium Services are a departure from the typical approach to library services and one that requires careful consideration before implementing. The potential Premium Services areas identified thus far have been divided into the following categories:

- **E-Conversion Services** – creation of digital content from non-digital materials and custom, on-demand transformation of materials from one format to another
- **Metadata Services** – needs assessments, feasibility studies, advice on appropriate taxonomies, metadata crosswalks, metadata creation and support services, etc.
- **Custom Repository Services** – expansion of standard DSpace storage allocations to meet Community or individual's requirements that exceed normal limits
- **User Reporting Services** – research alert services, targeted notification services, hot topic citations, and custom reporting services

4.0 Cost Model

DSpace differs from other digital library initiatives in that it captures content and descriptive metadata directly from the creators through a distributed web-based submission process. Self-defined subsets of the MIT academic and research community (such as schools, labs, centers, or departments) will determine for themselves what research materials may be submitted. This experimental approach is a shift in the way that Libraries have traditionally managed their information assets, giving faculty members more control over what materials are collected and how they are described and shared with readers. It also distributes a portion of the collection management costs to the research Communities.

Our cost model was developed to assist MIT Libraries in staff planning, understanding the budget necessary to sustain DSpace and the factors that may significantly affect our projections. Therefore it is necessarily a forward-looking analysis and is presented from the Libraries' perspective. In order to estimate costs we consider the changes that the implementation of DSpace has made in staff, support services, space and hardware requirements thus far and how we expect their impact on total costs will change with growth of the system.

Growth projections are based on current staffing, increasing only at 4% per year, the average wage and benefit increase over the last ten years at MIT. Because of our need to control costs carefully we estimate that we can serve the broad number of researchers at MIT with growth to 3TB storage in three years. Beyond three years we considered any projections on storage needs to be far too speculative. The Digital Libraries Research Group at MIT is continuing to develop projections and we look to other ongoing research as well.

We classified costs according to how they will be allocated to DSpace relative to the overall Libraries' budget, and present the FY2003 estimates.

Incremental	Costs that will create new expense categories: <ul style="list-style-type: none"> • Two new dedicated staff • Minimal operating expenses • System equipment escrow All Incremental costs represent actual cash flows.	\$215,000
Principal	<ul style="list-style-type: none"> • Prorated salary and benefits for existing library staff with an anticipated DSpace allocation greater than 20% • Additional operating expenses • Travel 	\$50,000
Comprehensive	<ul style="list-style-type: none"> • Prorated salary and benefits for staff with an anticipated DSpace allocation greater than 5% 	\$20,000
Total		\$285,000

Table 1. DSpace Cost Classifications

We then reported the allocation of Incremental, Principal and Comprehensive costs as applicable in traditional accounting categories: Staff, Operating Expense, and System Equipment.

Staff Salaries and Benefits	All affected staff at various allocations, including benefits	\$225,000
Operating Expenses	General office, PCs, travel, Information Systems SLA	\$25,000
System Equipment Escrow	Allows for growth to 3TB in 3 years at today's prices	\$35,000
Total		\$285,000

Table 2. DSpace Budget Allocations

4.1 Staff Salaries and Benefits

The Transition Planning Group members tasked with the staffing plan conducted interviews with a broad range of staff members throughout the Libraries and categorized the impact of DSpace on individual staff members from minimal to high. We used the results of the Transition Planning Group's staffing plan as input data for staffing costs, translating the impact levels to staff allocation percentages.

The Libraries recognize that the ability to foster and respond to rapid early growth effectively will be critical to the acceptance of DSpace by MIT Communities. Although the skills required to run DSpace exist among current Libraries staff, it was determined that relying exclusively on existing staff would provide a service far too fragmented for success and, thus, the Libraries determined that DSpace will be staffed with two new

dedicated staff, a DSpace User Support Manager and a DSpace Systems Manager. Costs for these positions, including staff salaries and benefits, office supplies and equipment are included in Incremental costs.

All other staff impacted by DSpace were categorized into either Principal or Comprehensive costs according to their estimated allocation. For example, the Associate Director for Technology and the Systems Office staff were allocated as Principal costs because their anticipated DSpace time allocation is greater than 20 percent. Public Service staff, who interact significantly with the Libraries' users and answer questions about the Libraries' electronic resources, were accounted for under Comprehensive costs, in the interest of understanding the full impact of DSpace on Libraries' planning and the contribution of the existing MIT Libraries infrastructure.

4.2 Operating Expenses

Incremental Operating Expenses include offices supplies and expenses for the dedicated staff and the Information Systems Service Level Agreement fees. Principal Operating Expenses include staff travel, conference fees and other meeting expenses.

4.3 System Equipment Escrow

We outlined growth scenarios to our development partners at HP and worked with an HP hardware expert and a representative from MIT's Information Systems organization who helped us develop alternatives for scaling the DSpace system. The cost estimates were based on HP hardware costs and are included in Incremental Expenses. Other vendors may have higher or lower costs associated with hardware to meet these same specifications.

MIT Libraries is well positioned to delay any major hardware purchases until the actual growth of the system is better understood. As we gather data on the usage of DSpace, the DSpace User Support Manager will be studying the demands on the DSpace system in terms of cumulative content. Based on our current best estimates of typical demand, the hardware donated by HP is more than sufficient to handle the anticipated transaction volume, is well configured for disaster recovery, and can scale to store approximately 3 terabytes (TB) of data with the purchase of additional disks. DSpace was architected with scalability in mind. The storage is decoupled from the web interactive services such that the two can be scaled independently. We do not anticipate growth in the application services; our cost estimates therefore reflect growth only in storage.

4.4 Federation Note

Our model reflects costs specific to MIT, where DSpace will be implemented as a full-scale digital repository in an organization with access to MIT's robust information systems architecture but with limited excess resources. However, DSpace is a fully scalable system, operable on an individual personal computer and DSpace code will be offered at no charge through an Open Source license. The costs that a federating

institution may incur in the implementation and operation of DSpace are a factor of the intended use of the system (e.g. a pre print document server or a multimedia content repository), current staff availability, institution size and research output, and information systems resources. For example, DSpace implemented as a pre-print server in a small university or department may add little or no costs in the Incremental category.

5.0 Funding Sources

MIT Libraries plans to offer DSpace Core Services free of charge to all registered Community members to encourage widespread adoption of the system among faculty and researchers. The DSpace content is provided free of charge to the public. This necessitates that MIT Libraries find sources other than user or subscription fees to support the ongoing operations of the DSpace service.

We propose a funding model, which is anticipated in our cost model, for MIT Libraries that is supported by a number of resources in the form of financial support or in-kind assistance:

1. **Institutional Support:** MIT Libraries intend to seek support from MIT in the Libraries' annual operating budget. The total annual cost of operating DSpace, including new storage to accommodate projected annual submissions, currently represents less than 2% of the Libraries annual budget. The DSpace service is an extension of the Libraries' traditional role of capturing the scholarly record of the Institute and provides a single managed solution that is more advantageous for faculty and research communities that currently have to develop and manage ad hoc solutions on their own such as individual web sites.
2. **Collaborative Development:** HP, MIT Libraries, the MIT Lab for Computer Science and the World Wide Web Consortium, have begun a three-year research project, SIMILE, using the DSpace platform and content as a test bed to explore interoperability of complex community-defined metadata schemas. Funding for this project came from invent@MIT, the HP-MIT Alliance, and applicable research results will be migrated into the operational DSpace system.
3. **Federation:** The DSpace Federation will leverage the distributed expertise of Federation partners and capture their enhancements (e.g. in the areas of digital collection management, preservation, cross-institute system interoperability and scholarly communications) as new system development under Open Source guidelines. Collaborative development across universities will allow the system to advance far more rapidly than would be possible in a single instantiation.
4. **Premium Services:** MIT Libraries reserve the right to charge for Premium Services as a means to meet extraordinary user requests while controlling the impact on the Libraries' scarce resources. Those services for which MIT Libraries are the sole source provider, such as Custom Repository Services, or clearly can establish added value over other competitive services offer an opportunity for cost recovery by the Libraries.

6.0 Future Plans

The Cambridge/MIT Institute, a cooperative venture between Cambridge University and MIT supported by the British government, is providing the funding for a series of seminars designed to explore the issues presented in this paper in detail. The series, offered to senior library managers of HE and FE institutions in the UK, will begin in September 2003 and is meant to foster further discussion and provide participants with all of the information necessary to complete actionable business plans for a digital institutional repository within their own institution, regardless of software platform chosen.

Additionally, CMI has provided funding for the first federated implementation of DSpace in the U.K. at Cambridge University. The project began in November 2002 and will focus on developing interoperability, supporting digital course materials, and using the digital content of DSpace as a test bed for digital preservation techniques.

MIT also has six Federation partners (Columbia University, Cornell University, Ohio State University, the University of Rochester, the University of Toronto, and the University of Washington) to be sponsored under a grant from the Andrew W. Mellon Foundation. Beginning in November 2002, these universities will install and evaluate the DSpace system in their university environments and work with MIT to develop Federation organizational, governance and business models for the long-term management and maintenance of the open source system for the benefit of all member institutions.

7.0 Conclusion and Next Steps

Strategic, economic, and organizational issues are significant challenges that must be considered for a successful implementation of an institutional repository. Institute support, strong library leadership, and business and operational planning conducted in parallel with the research and development process were of paramount importance. As DSpace matures and we gather user feedback and usage data, this business model will evolve. Our Federation partners also will contribute to our understanding of the costs of institutional repositories. We plan a series of seminars to be funded by the Cambridge/MIT Institute, and offered to universities and other cultural heritage institutions in the UK. The series is designed to provide the necessary framework to develop individual business plan to implement digital repositories in each participant's respective institution, regardless of platform. Further we hope to foster collaborative discussion regarding best practices in digital library development. All of these activities will be used to inform future modifications of our business model.

References

Fleishauer, Carol, Mary Barton, Margret Branschofsky, Dan Chudnov, Catherine Friedman, Keith Glavash, Carl Jones, Jeff Merriman, Sarah Mitchell, MacKenzie Smith, Megan Sniffin-Marinoff, Julie Harford Walker. "Report of the DSpace Transition

Planning Group to the MIT Libraries Steering Committee.” Cambridge, MA., 19 April 2002.

Hedstrom, Margaret. *Digital preservation: a time bomb for Digital Libraries*. Available from World Wide Web: (<http://www.uky.edu/~kiernan/DL/hedstrom.html>)

“Dublin Core Metadata Initiative Overview.” Available from World Wide Web: <http://dublincore.org/about/>

Author Information:

Mary R. Barton and Julie H. Walker were hired by MIT Libraries under an Andrew W. Mellon Foundation grant to create a sustainable business plan for DSpace. Barton holds a B.S. in Economics with a concentration in Finance from the Georgia Institute of Technology and earned her MBA in Finance from the MIT Sloan School of Management. Walker holds a B.S. in Management with a dual concentration in marketing and information systems from Rensselaer Polytechnic Institute and earned her MBA from the MIT Sloan School of Management.

For information about the CMI sponsored Seminar Series, contact Mary Barton, mbarton@mit.edu.

For more information about the implementation of DSpace at Cambridge University, contact Julie Walker, jhwalker@mit.edu.

For more information about the DSpace Federation project, contact MacKenzie Smith, kenzie@mit.edu.