

OSS Watch

DSpace: a case study in sustainability

Contents

- 1. [Brief description](#)
- 2. [Introduction](#)
- 3. [Project history](#)
- 4. [Growth and development](#)
- 5. [Structure and governance: building a collaboration infrastructure](#)
- 6. [Structure and governance: sustainability planning](#)
- 7. [Reflections and future](#)
- 8. [Further reading](#)
- 9. [Acknowledgements](#)

In the latter half of 2006, the Joint Information Systems Committee (JISC) commissioned a study via its Teaching and Learning committee to examine the issues surrounding sustainability of open source software. The resulting report drew together seven case studies of successful but very different open source projects and examined each project's sustainability model. Each of these case studies has been told from the point of view of the lead developer or one of the key personnel and gives a fascinating insight into the factors that have determined the success of each project. These case studies are now presented by OSS Watch as stand alone documents in a series.

This case study, examining the DSpace project, has been written by Julie H. Walker, MIT Libraries.

1. Brief description

DSpace is a digital repository software platform with broad functionality for the capture, management, preservation, and redistribution of digital scholarly research materials in a variety of formats, for a variety of purposes. It is typically used to provide an institutional repository for open access scholarly research, theses and learning objects, and as a preservation archive. MIT Libraries and HP Labs started to develop the software in November 2000 as part of the HP-MIT Alliance, an innovative research and development initiative between industry and academia. Two years later, they released DSpace 1.0 as open source digital repository software under the BSD licence. By September 2006, approximately 175 research institutions located in more than 35 countries had registered a live DSpace repository.

DSpace has been funded through a mixture of private research investment, grants, and donations of time and resources from participating institutions. Although MIT and HP Labs currently own the copyright for the DSpace software they are in the process of creating a separate DSpace non-

profit company and plan to transfer ownership of the copyright to the company. The intention is to also use the company to develop an independent funding stream to ensure the sustainability of the platform. This way, everyone will be able to enjoy the benefits of shared maintenance and development of the software, and the greater impact on the research and standards associated with digital content dissemination and preservation.

2. Introduction

The story of DSpace's development and release offers an interesting take on the progress of an open source project. Although the two original partners (HP Labs (UK) and MIT Libraries) undertook the initial development in a proprietary fashion, it was always envisaged that the code would be 'opened up' and an open source community would be encouraged. In addition, the model for DSpace is quite unusual: its stakeholders are mostly large educational organisations, not individuals, and domain experts, not programmers, decide features and functionality.

Since the software was originally released, a community made up of organisations that either use or develop the DSpace open source software has emerged. Investment in DSpace continues to grow. Adopters employ technical and functional experts to build and expand local DSpace services and pursue various research agendas pertaining to the DSpace platform. For example, adopters gather and manage, for deposit in DSpace, collections of digital material that require long-term stewardship; they provide capital outlays for server, back-up and storage hardware; and they offer technical support to the community and commit resources to develop the DSpace platform further.

3. Project history

For a few years prior to the instigation of the DSpace project, HP Labs and MIT Libraries had shared an interest in finding a solution to the problem of providing a long-term home for born-digital material. Together, they wanted to build a simple, functioning system in order to learn more about the problems associated with ingesting and managing content. At the same time, they wanted to use an open source approach to encourage adoption and enable researchers and developers to enhance the software. In November 2000, as part of the HP-MIT Alliance, they began a research and development collaboration to create a software system with broad functionality to capture, manage, preserve, and redistribute digital scholarly research materials. They invested a combined sum of roughly US\$4million over two years to cover the costs of hardware infrastructure, software development and project management resources, facilities, and domain expertise.

Two years later, in November 2002, they released DSpace 1.0 as open source software and this is where the transition of DSpace from a closed, sponsored research project to an open, community-based project began.

4. Growth and development

Not surprisingly, the transformation to an open, community-based project did not occur overnight, and for nearly a year and a half, MIT and HP continued as the primary developers and supporters of the project. During the time between November 2002 and March 2004, other institutions focused internally on their DSpace installations e.g. downloading the software, implementing it, collecting content and building services that fitted the needs of their organisations. They often asked for technical support on the DSpace discussion lists, which were created as a collective resource to answer questions from the community. MIT and HP answered nearly all of the questions in the beginning, but as other institutions progressed further with their DSpace installations they began to draw on their experiences and offer help in this effort.

In 2003, MIT obtained a one-year grant from the Andrew W. Mellon Foundation to promote the adoption of DSpace and study six research universities' experiences with creating their own DSpace institutional repository service model. Funding from the project also supported several pivotal meetings, e.g. an installation training class that assembled DSpace technical resources from participating institutions for the first time, and several meetings of domain experts to address the complexities of institutional repository policies, service models and business planning. These meetings were instrumental in forming the DSpace Federation open source community: the group of colleges and universities, research institutions, government agencies and NGOs, companies and private collectors that constitute the community of software users.

Later on, around 2005, for-profit companies, including Hewlett-Packard and UK-based BioMed Central, started to establish business models around the DSpace platform, providing services and support to the community and also contributing some platform enhancements. Commercial involvement with DSpace is a milestone for the project because it indicates the potential that companies see in the technology and signals that the platform has reached the level of stability on which a business can be built. It also increases DSpace's reach, in that many new, smaller groups of users can join DSpace without the need for the level of ICT resources that would normally be the preserve of larger organisations. However, there are also some disadvantages, such as issues of control and intellectual property, which need to be carefully balanced. As the stakes grow higher for those involved with DSpace, so too do the demands for more co-ordination and greater long-term stability for the project.

5. Structure and governance: building a collaboration infrastructure

MIT organised the final DSpace Federation project meeting, with funding from the Mellon grant, Hewlett-Packard and Google. An invitation to the meeting, held in March 2004 at MIT, was extended to all DSpace adopters, totalling approximately 50 institutions from 9 countries, and effectively became what was later known as the first User Group Meeting. It brought about a more formal development structure within the community and proved a milestone in the evolution of DSpace. The participants agreed they needed a better model for allowing users to modify the software. They looked to the Apache Software Foundation for a solution and decided to adopt Apache's 'committer group' approach, so called because specified individuals have rights to 'commit' (i.e. submit) new code to the software. The DSpace committer group began with five software developers who represented particularly active DSpace user institutions, and has since expanded its membership to nine developers, selected on the basis of their individual contributions to the software. These committers dedicate time to system architecture planning and development, feature planning and development, bug fixing, integrating new code submissions, quality assurance testing, release management, documentation, technical support, participating in specialised work groups, and many other tasks to maintain and enhance the platform. The more general pool of 'contributors' actively supports the community by answering technical questions from users and providing bug fixes and software enhancements. A set of guidelines defines the steps to contribute code to DSpace. With an established committer group and a clearer path to submit and integrate new code, the community gradually began to engage in the software development process.

A basic infrastructure supports the efforts of the developers. The DSpace developers deposit code centrally in SourceForge, the world's largest open source code repository, and use SourceForge's issue tracking database. SourceForge and MIT host group discussion e-mail lists such as 'DSpace-General', a general purpose list for non-technical discussions; 'DSpace-Tech', offering technical support; and 'DSpace-Devel', a discussion group for code and feature development and release. The 'DSpace wiki', which has been integrated with the main, informational DSpace website, provides a participatory forum for community members to update others on the latest DSpace news and developments. MIT and the committer group currently provide most of the upkeep for this infrastructure.

As the originating sponsors of the project, developers of the platform, and copyright owners of the DSpace software, MIT and HP Labs have provided a substantial amount of co-ordination and infrastructure over the years to foster the project's development. Even now, they still provide overall project guidance, technical resources, infrastructure, and general co-ordination of various aspects of the project. Other institutions are increasingly involved in these efforts, reflecting their growing investment in the platform and the health of DSpace's open source community.

6. Structure and governance: sustainability planning

In 2003, MIT and HP Labs, with input from the other DSpace adopters, began investigating the options for sustaining the DSpace software in a way that would not rely exclusively on the two founders and would embrace the open source software approach to managing software for and by the adopters. By the time of the second user group meeting, held in July 2005 at the University of Cambridge, UK, and funded in part by the Cambridge-MIT Institute (a joint venture between MIT and the University of Cambridge), the community had grown to over 80 user institutions in 22 countries. With the emergence of its thriving open source community, MIT and HP believed that DSpace had reached an appropriate point in its development at which it needed a representative governing and legal ownership structure that would reflect its status as a shared, open source software resource and provide for the long-term sustainability of the software. A session on governance at the July 2005 User Group Meeting confirmed this belief, as representatives from other institutions spoke out about the need for more central organisation and decision making in areas such as core development and support, roadmap planning, quality assurance, release management, documentation, and training. In addition, the group identified a need for a sustainable business model, legal guidance on intellectual property and liability, and a more robust collaboration infrastructure.

The community agreed to a governance planning process that called for the formation of an interim advisory board, made up of representatives from MIT, HP Labs, other DSpace installations, DSpace service providers, and experts in open source software and the majority domain of use (e.g. higher education and research libraries). The advisory board was tasked with drafting a recommendation to the community for a governance plan to support its activities.

With funding from the Cambridge-MIT Institute, the advisory board convened on March 30-31, 2006 in Cambridge, MA. At the meeting, the advisory board considered a variety of options: (1) starting an independent non-profit organisation; (2) joining an existing non-profit (i.e. another open source software organisation such as the Sakai Foundation or the Apache Foundation via its incubator programme or a related library or higher education organisation); and (3) starting or joining an existing for-profit (i.e. a commercial service organisation supporting the DSpace community). After considerable discussion and debate, the advisory board recommended that an independent non-profit organisation be formed. The board felt that this kind of organisation would provide the co-ordination needed and the best approach to continued innovation and development, while not precluding the possibility of creating or joining an umbrella organisation of some kind in the future.

The advisory board also addressed the fact that contributions from the developer community had focused primarily on extending end-user functionality, fixing bugs, and supporting users, which meant that the underlying architecture and core functionality had received little attention. These areas required a more substantial time commitment and, probably, dedicated resources. Early attempts within the committer group to reach an agreement on the technical approach had resulted in technical stand-offs, which remained unresolved for lack of facilitation or a method for charting a path forward. The board recommended a formal architecture and technology review process, directed by a committee comprised of DSpace developers from the open source community and experts in the digital library field. The architecture review process was envisioned as a way to overcome the obstacles the committer group had encountered and focus attention and resources on an aspect of the platform that was not getting addressed through a

more grassroots approach. The process included an analysis of necessary changes to the DSpace architecture, and a roadmap and set of specifications for the implementation of those changes.

7. Reflections and future

In the past six months since the advisory board's meeting, working committees have begun to execute the board's recommendations. The documents required by state and federal agencies in the U.S. to start a non-profit company have been completed and a Board of Directors has been selected. The board plans to file the paperwork to register the non-profit company by the end of March 2007 and begin operations shortly thereafter.

The initial goal was to start with a small board of appointed representatives from institutions with both a significant investment in DSpace and an ability to execute the necessary tasks needed to start up a non-profit company. The idea was that by the end of the first year of operations, the board could expand to as many as 11 members. The additional members could include a Managing Director (to be hired in Spring 2007) and a Chief Technology Officer (CTO) (to be hired sometime in the first year or two). The Managing Director would direct the daily operations of the DSpace company, including responsibility for membership and resources, marketing and community building, outreach to other projects and initiatives, legal oversight, liaison with DSpace service providers and fundraising. The CTO, if deemed necessary, would co-ordinate the platform development including core development and support, process facilitation, collaboration infrastructure, and general technical oversight. The remaining Director slots and any future vacated slots will be elected by the broader DSpace community, with the goal of gaining broader representation from the various types of institutions using DSpace or contributing to the DSpace open source software. Board members will be asked to serve three-year terms and the terms will be staggered in such a way that only a third of the Board should turn over in any given year.

Following on from the advisory board's March 2006 recommendation, a DSpace architecture and technology review process was established, and a group of 12 technology experts participated in the effort. The group began work in August and completed their report by the end of 2006. Documents were created to describe the areas of the architecture that require analysis, working principles for the group to adhere to, and the actual process the group should follow to accomplish their work. The Chair of the group had deep technical expertise and knowledge of the digital library domain, and this enabled him to facilitate the group's decision-making process. After a week-long workshop in October 2006 to review the known architectural issues, the group designed a new DSpace 2.0 architecture that was published, along with more detailed implementation plans, in December 2006.

A sustainable business model for the non-profit remains an open issue for the Board of Directors and Managing Director. MIT and HP Labs are committed to helping with the start-up costs of the organisation, but the non-profit will need to develop an independent funding stream beyond that.

Some of the costs for sustaining the community are shared, such as platform maintenance and development, user support, and documentation, which are contributed by staff at user institutions. Other centralized activities such as those managed by the Director and the CTO require funding.

Work on the sustainability model for the DSpace community is a work-in-progress and this case study only captures a snapshot in time. Generally speaking, it is fair to say that successful open source projects will need to react to growing pains and institute various more formal, structured means for managing their development. DSpace hit this point in mid-2005 and chose to put together an advisory board to recommend a plan for a more formal governance structure. What we had learned from other open source projects is that we shouldn't proactively put structure in place but should react to the needs voiced by the community. Grow organically, in other words.

With hindsight, some of the lessons learned from early difficulties with the technical approaches could have been ameliorated by just getting everyone together in one room to resolve different points of view. The architectural review process achieved a positive result by appointing a knowledgeable facilitator to help with this, and got a broad range of stakeholders from the community together to agree a detailed plan. However, the pace of change to the software and growth of the community is such that the model will continue to evolve for quite some time. Openness and collaboration are values that guide the project decisions and the community is willing to experiment with different models in an effort to best address its needs.

8. Further reading

Links:

- DSpace website [<http://www.dspace.org/>]
- JISC sustainability study [http://www.jisc.ac.uk/media/documents/programmes/distributed_elearning/sustainability_study-1.0.pdf]

Related information from OSS Watch:

- [Case studies index page](#)
- [Open source projects: criteria for success](#)
- [A guide to participating in an open source software community](#)
- [Crossing the Chasm: open source software comes of age - Open Source and Sustainability, Oxford, 10-12 April 2006 - conference report](#)

9. Acknowledgements

The sustainability study from which this case study is taken was commissioned by the JISC Learning and Teaching committee and funded from HEFCE's IT Infrastructure funds. The Learning and Teaching committee is responsible for supporting the learning and teaching community by helping institutions to promote innovation in the use of ICT to benefit learning and teaching, research and the management of institutions.

The sustainability study was edited by Gaynor Backhouse of [IntelligentContent](#) and her editorial guidance has contributed in large part to the excellent result.

Date: July 2007 (last updated Tue, 13 Nov 2007)

Author: Julie H. Walker, MIT Libraries

Unless otherwise indicated, this page is © 2007 [University of Oxford](#).

It is licensed under the [Creative Commons Attribution-ShareAlike 2.0 England & Wales licence](#).

OSS Watch is funded by the [Joint Information Systems Committee](#).

OSS Watch values your input and questions.

If you have feedback on this document, or any OSS Watch activity, please send it to info@