ScieCom info
Svenskt Resurscentrum
för vetenskaplig kommunikation
www.sciecom.org

DSpace Federation 2nd User Group Meeting

Jonas Gilbert, Web Coordinator for
the Digital Library, Göteborg University
Jonas.Gilbert@ub.gu.se

One of the most used tools world wide for Institutional Repositories is the DSpace software. It is today believed to be the most used platform for IR in the US, while EPrints still is the commonest platform in Europe. DSpace was developed by MIT and Hewlett-Packard and version 1.0 was released in November 2002 under an open source-license. While the initial development of the system was funded by MIT and HP, the subsequent development has depended on the efforts of the user community. The R&D section of Hewlett-Packard – HP Labs – still keeps a close engagement in the project and the system architect for DSpace, Robert Tansley, holds a position there. HP looks at this investment as a way to be leading in the area of digital preservation, and says that they will stay involved as long as the community shows a keen interest for the platform. MIT, as a user of DSpace, obviously has an interest in the progress of the system and the staff that works with DSpace at MIT has been very important to the whole user community by their efforts to coordinate websites, administer mailing lists and arranging user meetings.

A first user meeting was held in Boston in March 2004, and a second user meeting was held July 7-8 this summer in Cambridge, UK, and gathered 140 participants from 22 countries. A recurring theme in the presentations and discussions during the conference was the future development of the system, both from technical as well as functional and organizational aspects. Presentations were given by among others Clifford Lynch, Executive Director of the Coalition for Networked Information and Matthew Cockerill,

Director of Operations of BioMed Central.

While DSpace originally was designed as an IR platform for academic communities, there are today many examples showing how the software is employed in a wider variety of purposes. DSpace is for instance being used in connection to an extensive Chinese digitalization project concerning museum collections held at Chinese universities – in all more than 100 collections each calculated to require 2 terabyte data when digitalized. Another example, and from a totally different area, is found in Kansas where the federal government is using DSpace as a repository for the state documentation. Yet another example shows DSpace as a tool for e-book publishing, something which has been tried by The Australian National University. In India the authority responsible for the collection of statistical data is using DSpace as a storage tool. In UK DSpace is being used to e-publish the reports from a national research program within the social sciences, and in connection to this project the administrators have developed visualization tools for DSpace that allow for analyses of how the researchers from different universities are collaborating. This list of how DSpace is being used in different areas could be made longer, but the main area of use is of course in line with the original purpose for the system, i e as an Institutional Repository for universities. A question that was raised at the user meeting concerned how these different areas could be made to collaborate in the further development of DSpace? Should the thousand flowers bloom, or is it of importance that the system has a clearly defined roadmap?

To address this question about the future roadmap for DSpace, MIT and HP will appoint an interim board whose object will be to consider the future administration and organization for DSpace. A report from the interim board is due next spring, and one main issue is to describe the position of DSpace with regard to the broader development within digital preservation and e-publishing. This question was also addressed by the closing speaker Clifford Lynch, who pointed out that DSpace of course not is developed in isolation, but also that we today don't fully comprehend the environment and that we don't know how a repository system such as DSpace in the future will interact with e-learning platforms, publishing tools and other applications. Lynch also observed that Institutional Repositories today is a serious question on the agenda for many research institutions, and that approximately 80-85 % of all universities are implementing or planning to implement an IR. Along with this Lynch also noted that research funding bodies such as NIH increasingly are demanding that applications for research projects should include plans for data management.

Lynch also said that it is possible to see a difference between Europe and US with regard to how the IR systems are being implemented and used. In Europe the e-publishing of the

traditionally printed documents is in focus, while the universities in US are more interested in integrating other types of material in the repositories. This could be material such as research data, video or sound recordings, course catalogs, newsletter or other kinds of documentation related to education and research. Lynch did not consider either of these ways to use the repository as a “wrong” way, but said that we must keep in mind that the way the system is being used naturally will guide the strategies for the future development. Should an IR, for instance, be able to manipulate the data? Or should all items be just “big bags of bits” that leaves all processing of the data to the user? There are also several important issues left to solve concerning the long term preservation. For instance must a way to geographically distribute the archives by mirroring be achieved. Lynch also concluded that the best guarantee for an archive to be operational in the future is when it is being used daily now.

BioMed Central is using DSpace for their service Openrepository.com, which is an opportunity for organizations not willing to install and run a repository on their own. BioMed Central works as the service provider, hosting the repository and offering different related services for the customers. For this purpose BioMed Central has been active in developing new functionality for DSpace. In his presentation Matthew Cockerill from BioMed Central declared that these developments would be made accessible to the DSpace community.

A Nordic contribution on the user meeting came from the University of Bergen and librarian Elin Stangeland who described the plans for making a connection between the DSpace repository in Bergen and the Norwegian national system for research documentation, FRIDA.

It is very stimulating to follow the discussions about the future development of the DSpace platform. It is a good example of the possibilities and issues that comes with an open source project. If there sometimes in the discussions were a focus on problems, others would point out that these are natural questions that follow with a successful project. The need to co-ordinate the development of the system and the information and support is obvious. There also was a discussion about what it means to be a registered user of DSpace, and if there should be a more formalized possibility to enter the community of DSpace users. Many current DSpace users are already giving important contributions to the development. Among the examples of functionality that has been first locally developed and then integrated in the main software are: multi language support (University of Patras, Greece), authority lists for subject headings (University of Minho, Portugal), possibility for the readers to write comments to items in the archive (New York State University)… this list could be made longer and is a good illustration of the benefits that the users in the community can

have from each other in an open source project.

Svensk sammanfattning

Ett av de mest spridda programmen för Institutional Repositories är DSpace, utvecklat av MIT och Hewlett-Packard och lanserat i november 2002. I dag beräknas fler än 100 installationer av DSpace finnas i drift, spridda över hela världen. Många av dessa användare deltar nu aktivt i utvecklingen av DSpace, genom att skapa nya funktioner som sedan integreras i själva programvaran. Ett användarmöte hölls i somras i Cambridge och samlade 140 deltagare från 22 olika länder. Vid mötet diskuterades frågor kring den fortsatta förvaltningen av DSpace, och MIT och HP presenterade en plan för att tillsätta en interimstyrelse med uppdrag att utreda hur det framtida samarbetet kring DSpace ska organiseras. På användarmötet talade bland andra Clifford Lynch från Coalition for Networked Information som konstaterade att Institutional Repositories är en viktig fråga på många lärosätens agenda, och att 80-85 % av universiteten idag antingen driver eller planerar att driva ett institutionellt arkiv. Samtidigt blir det också allt vanligare att forskningsfinansiärerna ställer krav på att ansökningar ska innehålla planer för data management.