

# Electronic publishing and institutional archives – utilising open-source software

Contribution to the ELAG conference, Barcelona 9-11 May 2007.  
Ellen Røyneberg, BIBSYS

## 1 Introduction and background

The interest for open access and institutional archives in Norway is growing. In 2005 four of universities already had an institutional archive or specific plans for developing such systems. In the last quarter of 2005 the remaining university libraries, some university college libraries, research libraries and BIBSYS started to discuss the possibilities of a joint effort to develop institutional archives. During these discussions the participants agreed to initiate the Pepia project (Project for Electronic Publishing and Institutional Archives). The objective of Pepia was to create institutional archives for the participating institutions.

BIBSYS has more than 30 years of experience in the field of non-profit consortia solutions for library automation and was tasked to develop the software in close cooperation with the consortium. The project received 50% funding from the Norwegian Archive, Library and Museum Authority, the remainder was funded by the participating institutions.

### 1.1 *Pepia*

The Pepia project started in January 2006 with a project group containing members from three of the consortium institutions and a developer and project manager from BIBSYS. The group created a requirement specification for an institutional archive for the consortium. In addition, they made a survey of relevant systems that could be used as software platform. Furthermore, the project group evaluated five systems, DSpace, EPrints, Fedora, Diva and LOCKSS and recommended DSpace as the preferred software platform.

Based on this recommendation, BIBSYS customized DSpace to meet the requirement specification. The new customized DSpace version called BIBSYS Brage was launched as a beta version in December 2006. The project is considered a success but there is still some upgrades needed. BIBSYS has received fresh funding from both the consortium and Norwegian Archive, Library and Museum Authority to further upgrade the system. It is expected that the upgrades to further improve the system will be implemented during 2007.

## 2 Utilising the DSpace software

The most important reasons for the project group to choose DSpace as software platform are described below:

- DSpace fulfilled more of the requirements than its competitors
- A number of large universities worldwide have experience with DSpace
- DSpace can be hosted on an existing server platform at BIBSYS. This meant that BIBSYS did not have to acquire new competence
- DSpace is an open source software (OSS)

The goal of the Pepia project was to give the participating institutions an institutional archive. An

institutional archive is said to be the green route to achieve open access (OA). The idea behind OA is to provide free and unrestricted access to scholarly material, primarily peer-reviewed research articles. It would therefore be a paradox to use proprietary systems when developing a system with the objective for free access. (Jones, R 2006) (Jones, R and Andrew, T 2005)

## **2.1 Installation**

A standard DSpace installation runs on a Tomcat servlet container and with a connection to a PostgreSQL database. BIBSYS use WebSphere Application Server and it was therefore necessary to configure DSpace to get the application to run successfully on the server.

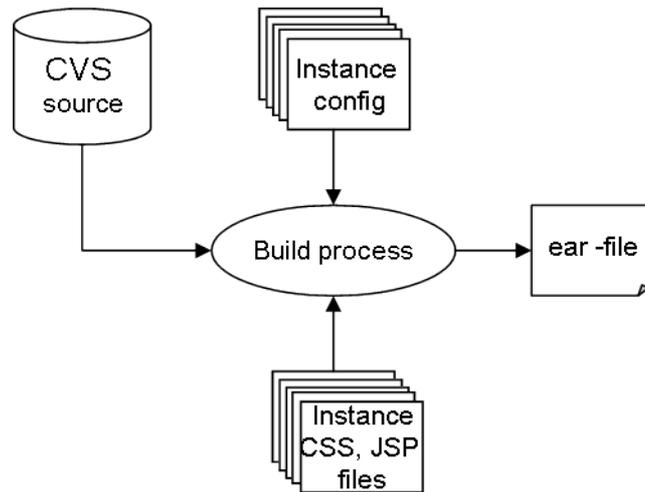
Initially, DSpace with out of the box functionality was installed. This first installation proved to be time consuming, but the results were good. However, the DSpace development structure was not compatible with the BIBSYS integrated development environment (IDE). At this early stage in the project, the project team had limited familiarity with DSpace and the installation guide did not provide the necessary details to anticipate this issue.

Consequently a substantial amount of time was used to understand the DSpace development structure and build scripts. In addition, the details of the BIBSYS IDE was mapped to be able to identify possible solutions. The final solution was to change the DSpace source directory to fit the IDE structure and to tailor the IDE build script to create the DSpace application layout. This made it possible to deploy the application locally on the IDE server.

## **2.2 DSpace in a consortium model**

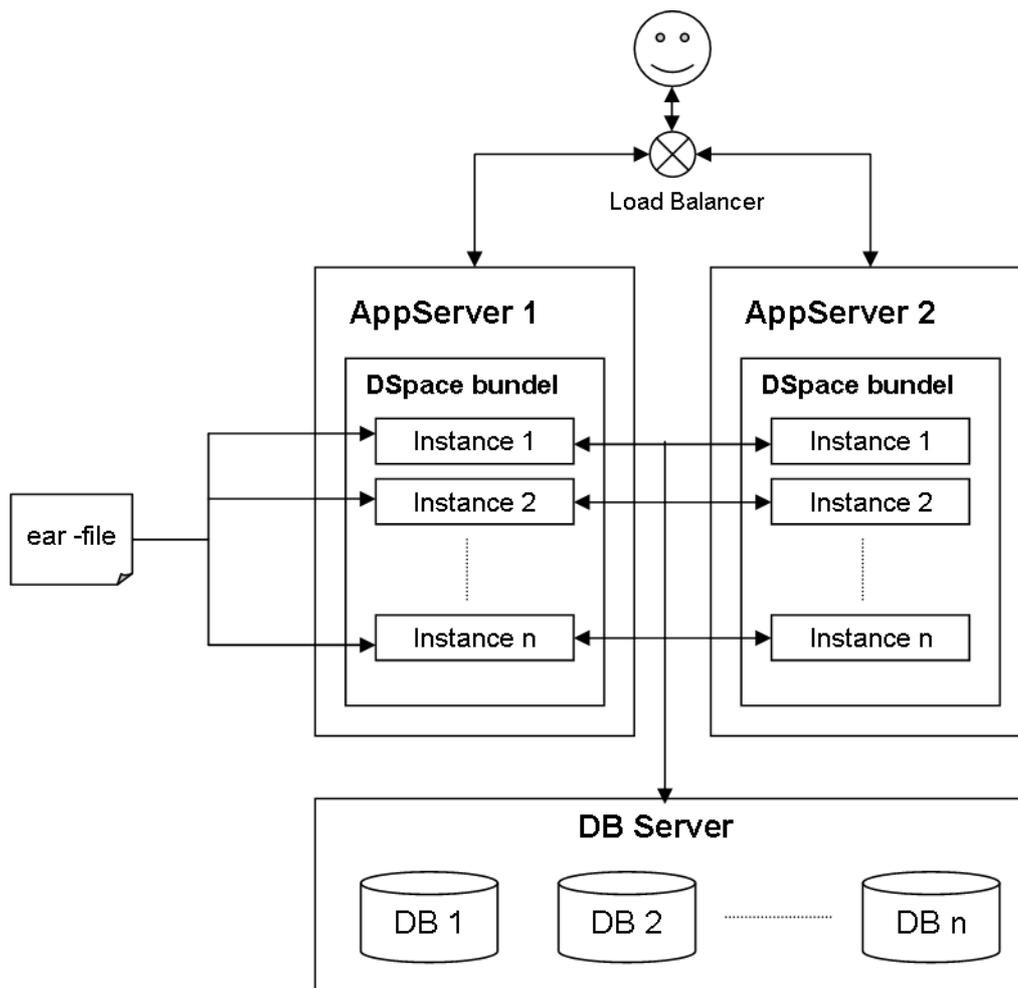
As mentioned the participants in the Pepia project is organised as a consortium. The initial plan was to make a joint system, with a common database and user interface for all the participating institutions. As the project gained more information about institutional archives the importance of providing each institution with its own separate archive and database became apparent. This required a development of an environment which could easily create more than 30 applications, while editing one source code. Furthermore it is also envisaged that the system should be able to offer individual customized user interfaces. Consequently, a method of deploying all instances at once was required.

Illustration 1 describes the new DSpace building process used in BIBSYS Brage. This build is separate from the IDE, but checks out the source that has been committed to CVS from IDE. With this build all the applications use the same source code but they can have separate configuration, css and jsp files. This process allows development and maintenance of a single source code only while providing the participating institutions the ability to customize the graphical user interface of their individual archives.



*Illustration 1: The new DSpace building process.*

As shown on Illustration 1 the result of the building process is an ear-file. The ear-file contains one war-file per application and is deployed on to the WebSphere Application Server. This server then unpacks the files and all the applications are simultaneously deployed. The applications are deployed on mirrored application servers with a load balancer routing the traffic as shown in Illustration 2. (Joki S, 2006) All the instances have a database connection, the database server is also mirrored for server stability.



*Illustration 2: DSpace on the application servers.*

These changes have provided an efficient work environment, however the project used more time and resources than anticipated which lead to less time available to adjust the DSpace functionality.

### **2.3 Adjusting DSpace functionality**

DSpace is a large and complex system, and since it is an OSS it is developed as a collaboration within the DSpace community. One of the benefits of choosing DSpace is the active and including community. The fact that a developer at any time has access to community experts has proven to be very powerful.

However, there are also some disadvantages with an OSS. Since it is a collaboration the system appears to have a high complexity. The code has different styles and the degree and quality of the documentation varies. The threshold of effectively using the system may appear high, but when the developer becomes familiar with the underlying concepts, DSpace is seen as a good foundation.

After having spent some time on understanding the system, the following functionalities were added in just a couple of months:

- Added ability for a submitter to “hide” an item for a specified period
- Integration with the registration system of the library
- The possibility to add more relevant metadata of a published article
- Translating the interface and help texts from English to Norwegian (the project received translations from two Norwegian universities already using DSpace)
- Customizing the user interface with a BIBSYS layout
- Changes in the user management

The user management in DSpace did not meet the project requirements and had to be modified. User management is always a complex matter, especially when trying to make a general system that can meet the requirements of several different institutions. The user management in DSpace is relatively simple; a new user has to register to get a user account. After the user account is created an administrator manually gives the new user access to collections of items. If the institution has a LDAP server, it can be used. All users in LDAP have access to DSpace, there is no need for registration. However, the user account is created at first time log in, and the administrator still needs to manually give the users access privileges.

Neither of these solutions would fulfil our requirements, as the biggest issue for large institutions is the authorization process. The process described above would require the institutions to relatively large administrative resources on managing user access to the collections. In order to rationalize an organization-rule functionality allowing administrators to set up access rules for given collections was created. This is achieved by creating a rule specifying which departments, faculties etc that should have access to individual collection. For this to work BIBSYS needs to know what department a user belongs to. BIBSYS currently does not have this information, but it will be made available by integrating the Norwegian access system, Feide, into DSpace. Feide allows application providers (like BIBSYS) to forward the authentication process to another system. In return BIBSYS get the necessary credentials.

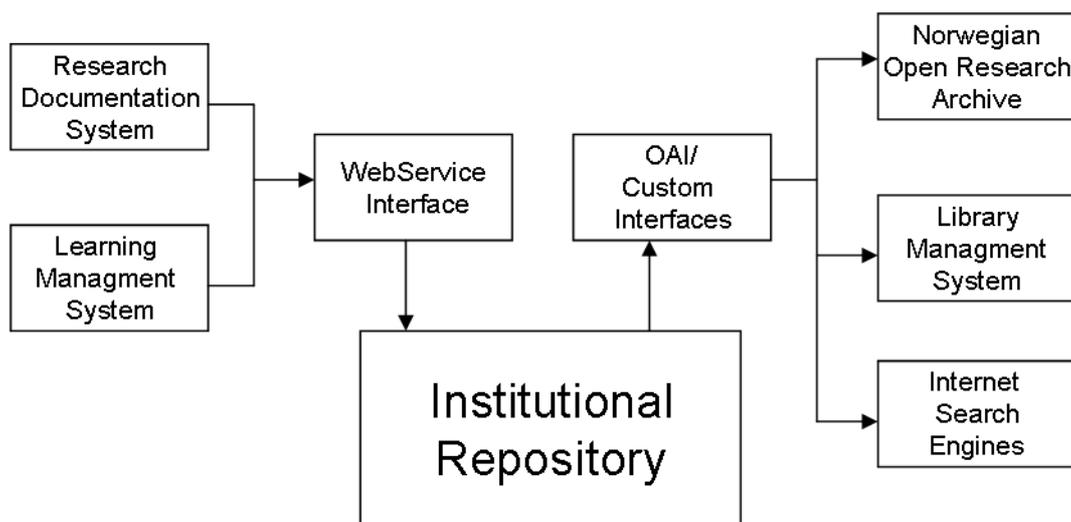
With the combination of the organization-rule functionality and DSpace authorization BIBSYS can provide a user-friendly, yet powerful way to manage users access to collections, with only minor changes to the original DSpace source.

### 3 Future development

As mentioned, BIBSYS will continue the development of BIBSYS Brage with the extra funding from the Norwegian Archive, Library and Museum Authority and the consortium this year. The current plan is to release a new version of BIBSYS Brage during first half of 2007 which includes the URN service form the Norwegian National Library and an integration to the Norwegian Open Research Archive, NORA. In the second half of 2007 we plan to start implementation of new functionalities however, the project scope is still being developed. Some of the features that will be considered to make BIBSYS Brage an even better product are:

- Integration with research documentation systems
- Predefined searches
- Integration with learning management systems
- Statistics
- Individual user interfaces for the participating institutions

Longer term, BIBSYS intends to create an archive which can easily communicate with or be integrated with other systems. Illustration 3 describes the envisaged integration of institutional archives to information management systems and output systems (Joki S, 2006).



*Illustration 3: The envisaged institutional repository*

### 4 BIBSYS and open source software

Traditionally BIBSYS has developed all its systems in-house. The funding of the Pepia project however, did not allow this and alternative solutions, like using OSS was required. Using OSS was a new experience for BIBSYS, but has proved to be interesting. BIBSYS intends to become more active in the DSpace communities.

BIBSYS is now in a process to further develop and provide our users a more modern library system. At the time of writing it is not decided if this will be done by develop the system in-house, or buy acquiring commercially available systems. In either case the experiences gained by using OSS will be valuable in the modernisation process. It may also be interesting to investigate the possibility of licensing some of BIBSYS systems (or parts of systems) as open source.

## 5 Conclusion

The Pepia project is the first project in BIBSYS where OSS has been utilised and we are very pleased with the results. Out of the box DSpace had most of the necessary functionalities that an institutional archive needs. Only a few functionalities had to be altered or added and it would have been impossible to reach similar results with the available resources without DSpace.

Our experience with OSS is mostly positive, but there are some disadvantages that one needs to be aware of. OSS are often intended for a specific task, application platform etc. It can be time consuming to tailor such fundamental elements, and it may be difficult to estimate the resource requirements in the scoping phase. Active use of the user groups in the communities are very valuable in all phases of a project.

With the combination of the expertise in the DSpace communities and our own experience we are confident that BIBSYS Brage will become a system that will be useful for the consortium members and for open access in Norway.

## 6 References

Jones, R and Andrew, T (2005) "Open access, open source and e-theses: the development of the Edinburgh Research Archive", Program, 39 (3), 198-212

Jones, R (2006) "Open Access Open Source and Institutional Repository" Internet Librarian, London 2006

Joki, S (2006) "PEPIA: A Norwegian collaborative effort for institutional repositories" OCLC Systems & Services (in print)

DSpace System Documentation: Installation (<http://www.dspace.org/technology/system-docs/install.html>)