# Fedora Commons 3.0 versus DSpace 1.5 : Selecting an enterprise-grade repository system for FAO of the United Nations

*An extensive evaluation of the Fedora Commons 3.0 and DSpace 1.5 digital document repository systems has been conducted. The evaluation aimed at selecting an open source software package that best satisfies the FAO Open Archive and FAO organizational requirements and the requirements for the storage, dissemination and preservation of documents and bibliographic metadata. Both repository systems were evaluated against thirty-two criteria chosen from nine core categories of requirements: community, security, functionality, integration, modularity, metadata, statistics and reports, preservation, and outputs. These criteria were selected with the merger of the FAODOC and FAO Corporate Document Repository (CDR) into the FAO Open Archive in mind.*

## Introduction

The Food and Agriculture Organization of the United Nations (FAO) maintains a number of heterogeneous document and document metadata repositories. The FAO Online Catalogue (FAOBIB)[1] is the online catalogue for documents and publications produced by FAO since 1945, non-FAO material added to the library since 1976, and serials held in the FAO library. FAOBIB catalogues and indexes both electronic and printed documents. The three FAOBIB collections are managed by three different subsystems: FAODOC[2], for FAO material; FAOLIB[3] for non-FAO material acquired by the Library and SERIALS[4] for serials records. The FAOBIB records have been totally created by information management specialists (cataloguers) and contain high quality descriptive metadata.

The FAO Corporate Document Repository (CDR)[5] contains full-text publications produced by FAO technical departments. The CDR[6] disseminates full text documents and a minimal set of metadata. The CDR uses a workflow system based on Electronic Information Management System (EIMS) to collect metadata through the course of the document/publication production process. The objective of EIMS is to have authors or producers of documents delivering the necessary administrative and descriptive metadata.

There is a lack of integration within the different bibliographical metadata repositories and the overlapping at content level unfortunately implies some inconsistencies that may affect the proper dissemination of the FAO publications. In addition, KCE duplicates efforts in cataloguing and maintaining technically different systems.

The objective of merging the FAO Corporate Document Repository (CDR) and the Catalogue for Documents produced by FAO (FAODOC) is the creation of one sustainable digital repository offering a solid foundation for the collection, management, maintenance and timely dissemination of material published by FAO. To improve the effectiveness of the proposed repository it will be necessary to streamline the current workflow and to integrate new modules. With the establishment of this digital repository FAO will take an important step in promoting the Open Access Publishing model within the food and agriculture community.

---

[1] The FAOBIB uses a CDS/ISIS OPAC at http://www4.fao.org/faobib/
[2] The FAODOC contains 166 482 records
[3] The FAOLIB, 72 693 records
[4] The SERIALS, 14 005 records
[5] The CDR is the gateway to FAO's publications at http://www.fao.org/documents/
[6] The CDR holds more than 28 000 documents

An extensive evaluation of the Fedora Commons 3.0 and DSpace 1.5 digital document repository systems has been conducted. The evaluation aimed at selecting an open source software package that best satisfies the FAO Open Archive and FAO organizational requirements and the requirements for the storage, dissemination and preservation of documents and bibliographic metadata. Both repository systems were evaluated against thirty-two criteria chosen from nine core categories of requirements: community, security, functionality, integration, modularity, metadata, statistics and reports, preservation, and outputs. All criteria were evaluated for each system by at least three members of the project team.

The objectives of the technical evaluation process were:
- To gain understanding of the design, architecture and implementation details of the two softwares.
- To evaluate the two softwares against an agreed set of criteria.
- To pay particular attention to the long-term development and maintenance lifespan of the softwares.
- To recommend the most suitable candidate open source software.

## Repositories chosen to be evaluated

Among the existing open source solutions the choice was restricted to Dspace and Fedora Commons tools, which were considered worth being evaluated because of their scope, longevity, stability and extended community of users as well as developers.

Fedora Commons is a Linux-based operating system that was designed in order to serve as a digital repository for a variety of use cases, including institutional repositories. The system offers a series of services and tools supporting the long term preservation of digital objects, the content versioning and the management of distributed repositories. More information is available at http://www.fedora.info .

DSpace is a digital repository management system; it was conceived for the setting up and management of institutional repositories and for digital archiving needs. A particular focus has been put in the subject of electronic objects management, digital preservation and publishing. It is Open Source and freely available, It is possible to customize it and extend it. It is backend by a solid community of DSpace users and developers. More information is available at http://www.dspace.org/ and http://wiki.dspace.org/.

## Evaluation Criteria

There are different criteria that are valid for all types of open source tools and provide a good indication of the software. Various reports are available on line related to the definition of criteria for the evaluation of software - open source or not.

The general criteria chosen for this scope have been mostly taken from a study carried out in 2006, which reviewed and evaluated all criteria found in open source literature. Under each of these general principles there is a number of detailed criteria that have been defined ad-hoc or extrapolated from literature. The general criteria selected for this evaluation were:
- Community
- Security
- Functionality
- Integration
- Modularity
- Metadata
- Statistics and Reports

- Preservation
- Outputs

Each selected criteria was given an importance rating to be used when evaluating the different Repository systems. Major criteria were also broken down into sub-criteria with each sub-criteria also having an importance rating. In order to have an objective and comparable evaluation each criterion has been assigned a weight, according to their flag: mandatory (***), important (**) or nice to have (*). A value ranging from 0 to 10 can be assigned:
- 10: Excellent support and meets the criteria out of the box, minimal effort.
- 8-9: Very good support with lacking features and needs a minimal amount of effort.
- 6-7: Good support but needs minor modifications.
- 5: Fair support but needs modification to reach the desired level of support.
- 3-4: Can be done but with significant effort.
- 2-1: Has poor support.
- 0: Feature does not exist.

## Evaluation of Fedora Commons

Fedora Commons supports the implementation of FAO Open Archive requirements. Its greatest strengths lie in its modularity and flexibility, in its capability of integration with other systems and databases, in its fine grained management of information and security. That, together with the extensive SOAP and REST webservices, makes Fedora Commons a very flexible and programmer-friendly system. The metadata schema is extensible without restrictions, while in DSpace the metadata schema is restricted at the collection level. Last but not least, Fedora Commons allows the creation of relationships between objects and allows the management of analytics and language versions of documents.

Fedora Commons is a complex system that requires good technological skill for installation and management but, although it does not have a user interface, can be easily extended. Choosing Fedora Commons grants the possibility to develop – and release – different modules at different times according to the strategic decisions of the FAO Open Archive Project Board.

It is relevant to point out that Fedora Commons ranks higher on those criteria that are most important for the FAO requirements:

- **Integration**. Fedora Commons uses a native XML representation of everything and extensive SOAP and REST webservices. It is extremely flexible and can easily interoperate with the existing and/or new FAO systems.
- **Preservation**. Fedora Commons fully supports the versioning of metadata and binary objects.
- **Modularity**. Fedora Commons is designed to work as a part of an extensible framework of service components.
- **Enriched metadata management.** Fedora Commons can easily manage any metadata schema, DC is the basic metadata schema but it can be extended without limitations.
- **Security.** The system uses XACML which provides a very flexible language for expressing access control policies with a fine grained approach.

## Evaluation of DSpace

DSpace strength lies in its being a complete solution and an easy to use system. The installation is well described and can be done quickly. It allows one to get started in just a few days and this is the reason why it is extremely popular, used all over the world and why it has a strong community of users and developers. However, detailed configurations and

customizations require more effort and time. Although the full source code is available, the extensibility is limited. Our requirements are numerous and complex and the level of flexibility provided by DSpace is not sufficient to support them. The approach of implementing add-ons could be a winning option only to handle some of the requirements, for most of them considerable changes to the source code are necessary.

In addition DSpace is a monolingually-biased system, a particularly inconvenient issue for international organizations like FAO with several official languages. While language attributes are (optionally) stored for all metadata elements, it isn't possible to customize the user interface using them. In the UI, you only have access to the elements themselves, and not the language information. So, for example, it is difficult to customize the Browsing views to display titles and abstracts in the language of the user's browser locale. DSpace seems to be currently in a state of flux. Many of the internationalization features are half-implemented, as well as many of the multiple metadata schema features. In conclusion, considering that the new release 2.0 should be available by 2009, the customization of DSpace at the core level is not the recommended option.

Dspace ranks higher on only 6 out of the 32 criteria and only 4 out of the 20 mandatory criteria. Interestingly these criteria are directly or indirectly related with the fact that Dspace has a ready-to-use user interface:

- **Strength of user community**. DSpace is an out-of-the-box solution, with a reasonable level of flexibility and it can be easily adopted and used as a solution for a repository management, this explains its wide distribution and user community strength.
- **User interface and workflow**. DSpace has a user interface and workflow with a certain level of customizability. But in any case, DSpace workflow is definitely not flexible enough to meet FAO Open Archive needs, so what in general might be a strength of the system is, in fact, one of its biggest limitations for the implementation of the FAO Open Archive.

## Conclusions

DSpace and Fedora Commons have been evaluated against 32 selected criteria, 20 of which were considered mandatory. Fedora Commons ranks higher on 21 out of the 32 criteria and ranks higher on 12 out of the 20 mandatory criteria. The two systems tied on 5 of the remaining criteria. As evidenced Fig. 5, Fedora ranks higher on most of the evaluated criteria.

The evaluation scores are summarized in the matrix available in *Annex 1.* The values provided by all participants have been normalized in order to harmonize them and make the comparable across evaluators: normalization was done by dividing each evaluator's Fedora ranking by their DSpace ranking. The evaluation matrix displays the average of all normalized scores.
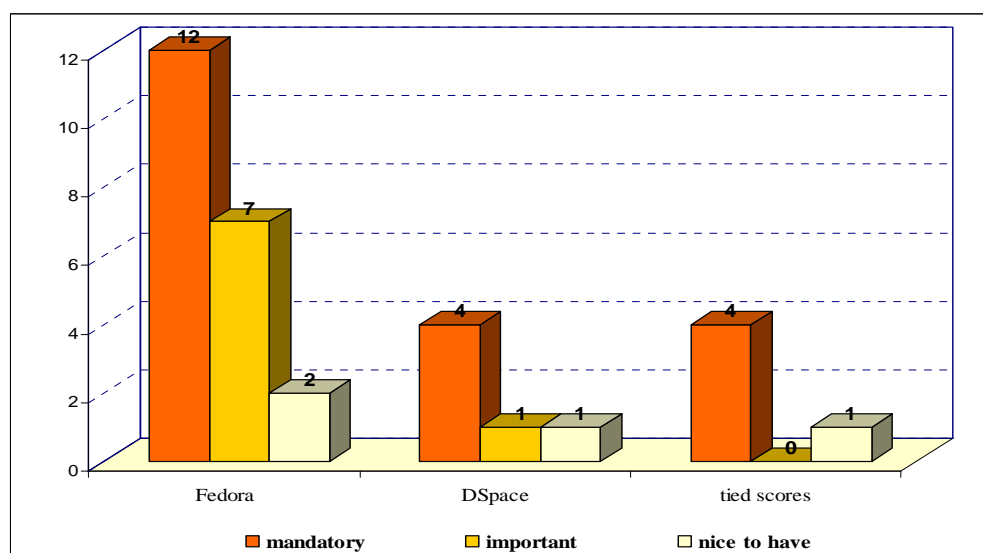
*Fig.1. Analysis of the evaluation matrix scores.*

Fedora Commons emerged through this evaluation as clearly the best candidate with respect to the project and organizational requirements for document publishing and cataloging. Fedora Commons was rated higher than, or equal to, DSpace on twenty-six (26) of the thirty-two (32) criteria. Moreover, Fedora Commons was consistently rated higher on the most important requirement categories: functionality security, integration and metadata. Fedora Commons was determined to be the more stable, modular and better documented of the two systems.

DSpace is a system targeted towards small, homogeneous document repositories for small-to-medium size organizations. As such, it does not offer the level of metadata representation flexibility, scalability, or modularity required by FAO document repositories. Fedora Commons is a stable, scalable and modular suite of webservices delivering complete, advanced and highly customizable digital document repository services.

The modular, service-oriented architecture of Fedora Commons will support incremental merging of FAODOC and CDR, as well as incremental deployment of advanced features to the FAO user communities. A staged implementation plan, in that it offers a high degree of flexibility throughout the project lifecycle, mitigates many of the risks involved in developing, deploying and maintaining the FAO Open Archive. Such a staged, incremental approach to the implementation of the FAO Open Archive is simply not afforded by the monolithic architecture of DSpace.

The recommendation of the Evaluation Team was to use Fedora Commons to implement the FAO Open Archive using an incremental, staged approach. This approach, which the modular, service-oriented architecture of Fedora Commons directly supports, will result in fewer service disruptions and minimize user disgruntlement with evolving features and user interfaces, while ensuring progress toward the end-goal of realizing a merged and consolidated document publishing and cataloging workflows in FAO.

## Bibliography

Evaluation of an Open-Source Repository System
http://www.switch.ch/de/els/LOR/evaluation.html. Last accessed in January 2008

Method for Qualification and Selection of Open Source software (QSOS) - version 1.6. http://www.qsos.org/download/qsos-1.6-en.pdf . Last accessed in January 2008

Karin van den Berg. *Finding Open options. An Open Source software evaluation model with a case study on Course Management Systems.* http://www.karinvandenberg.nl/Thesis.pdf. Last accessed in January 2008

Technical Evaluation of selected Open Source Repository Solutions
https://eduforge.org/docman/view.php/131/1062/Repository%20Evaluation%20Document.pdf . Last accessed in January 2008