eifl

KNOWLEDGE WITHOUT BOUNDARIES

HOW TO MAKE YOUR OA REPOSITORY WORK REALLY WELL

Checklist, based on EIFL webinars, to help you get the best out of your DSpace open access repository

Iryna Kuchma, Open Access Programme Manager, EIFL

December 2017

Table of Contents

Introduction

One of the EIFL Open Access Programme goals for 2016 – 2017 was to enhance open access repositories in EIFL partner countries. This means ensuring that they work well together with other systems and platforms, and include new repository functionalities that make repositories more user-friendly and enable easier sharing of research outputs.

We have therefore created this checklist of tips for improving open access repositories that are built with DSpace software. The tips are drawn from a series of seven webinars organized by EIFL with invited experts from the Institute of Development Studies (IDS, United Kingdom) and Stellenbosch University (South Africa), from January to May 2016. The experts, Nason Bimbe (IDS) and Hilton Gibson (Stellenbosch University), shared suggestions and good practices in setting up and running open access repositories using DSpace free and open source software, which is the most commonly used repository software in EIFL partner countries.

The first version of the checklist was published as a blog post on the EIFL website in June 2016. This is the second revised version. In addition to updates, such as recommendations for more recent DSpace software versions, we include two new sections. The first new section is on *Open Peer Review Modules for repositories*, drawn from an EIFL webinar organized in partnership with COAR, OpenAIRE and DIGITAL.CSIC and led by Isabel Bernal (DIGITAL.CSIC manager); the second new section is on *Next generation repositories*.

☑ WHICH VERSION OF THE DSPACE SOFTWARE DO YOU USE?

Always run your repository with the latest software versions (or no more than one version behind). The current latest release is DSpace 6.2 that can be downloaded from GitHub (dspace-6.2 release). DSpace 6.2 is a bug fix release to the DSpace 6.x platform. See the DSpace 6.x Release Notes for more information. Documentation is available at DSpace 6.x Documentation.

Find out more about how to upgrade your open access repository to the newer version of DSpace in this EIFL webinar and PowerPoint presentation.

If you want to upgrade your DSpace from 5.x to 6.x with XMLUI, Nason Bimbe has prepared useful notes, available here.

Also, you might like to read this article: Bimbe, N.B.; Lungu, S.; Kakana, F.; Sichilima, C.; Makondo, F.N.S. and Kanyengo, C.W. (2017): Challenges in Reinvigorating and Upgrading DSpace-based Institutional Repositories: A University of Zambia (UNZA) Library Case Study, IDS Working Paper 483, Brighton.

☑ HAVE YOU ENABLED THE HANDLE SERVICE?

Make sure you enable the handle service, which allows you to apply a short URL, which is persistent, for the purposes of citation and discovery on the web.

Find out more about enabling the handle service by visiting these websites:
http://hdl.handle.net;
http://www.handle.net/documentation.html; and
http://wiki.lib.sun.ac.za/index.php/SUNScholar/Handle_Server.

☑ DO YOU HAVE BACKUP/RESTORE PROCEDURES AND POLICIES FOR DISASTER RECOVERY?

Since your repository is now the vehicle for maintaining the permanent digital academic research record of your institution, you will be concerned about its sustainability. You will want to make sure it is backed up and monitored correctly.

Find out more about backup and recovery strategies and systems here.

☑ HAVE YOU ENABLED THE OAI-PMH SERVER?

So that your repository works well with other systems and platforms, it must be interoperable.

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a low-barrier mechanism for repository interoperability.

Data Providers are repositories that expose structured metadata via OAI-PMH. Service Providers then make OAI-PMH service requests to harvest that metadata.

Configurations for the OAI-PMH server are held in OAI-PMH crosswalks and [dspacesource]/config/modules/oai.cfg.] NOTE: unless you need to change the default behaviour, you may not need to alter OAI-PMH crosswalks and [dspacesource]/config/modules/oai.cfg].

To enable DSpace's OAI-PMH server, just make sure the [dspace]/webapps/oai/ web application is available from your Servlet Container (usually Tomcat). You can test that it is working by sending a request to: http://[full-URL-to-OAI-PMH]/request?verb=Identify.

If you are harvesting content (bitstreams and metadata) from an external DSpace installation via OAI-PMH & OAI-ORE, you should first verify that the external DSpace installation allows for OAI-ORE harvesting. More details are at https://wiki.duraspace.org/display/DSDOC6x/OAI and http://wiki.lib.sun.ac.za/index.php/SUNScholar/Remote_Harvest.

You can also read Peter Otuoma's guide DSpace – How to enable and configure OAI-PMH.

☑ HOW TO IMPROVE DISCOVERABILITY THROUGH SEARCH ENGINES

Search engines use software programmes that gather information from websites for indexing. These software programmes are called crawlers or bots. They are good at finding HTML files but since DSpace content is in a database system, you will need to tell the crawler how to get to that content.

First you need to make sure that you generate sitemaps. These can be generated through the Cron job # Generate sitemaps at 6:00 am local time each day 0 6 * * * [dspace]/bin/dspace generate-sitemaps. This will generate the sitemaps that are accessible at http://{your-DSpace-URL}/sitemap and http://{yourDSpace-URL}/htmlmap

Make sure also that the robots.txt contains directives to these paths. You will need the robots.txt placed in the root of your DSpace site. Make sure the robots.txt contains directives for what can and cannot be indexed Details of robots.txt structure and instructions are at https://wiki.duraspace.org/display/DSDOC6x/Search+Engine+Optimization.

Register your repository with Google Scholar and all known reputable registries and aggregators: OpenDOAR, ROAR, Base Search, CORE.

There is more information in this useful resource.

☑ DO YOU TRACK AND ANALYZE STATISTICS AND WEB TRAFFIC?

Having good statistics about use of your repository is useful in securing sustainability, for example, statistics can be used in advocacy, to increase buy-in and to build a business case to convince funders that your repository is worth supporting.

At the very least you should gather statistics about:
- Number of downloads and views;
- Where these are coming from (geo-location); and
- Content analysis, for example, how many items are in the repository (by various dimensions i.e. by item type, subject, language, etc.).

Assess what sort of insights you want from your DSpace and setup/acquire the appropriate visualization and analysis tools.

DSpace provides statistics such as the number of downloads, number of items, number of failed logins, etc. (all the data is available in the SOLR indexes – these can be queried directly as well).

Use Google Analytics (registration required) or Piwik (an open source web analytics) to collect information about site visits. Standard Google Analytics (JavaScript based) has limitations in capturing downloads.

In DSpace 5 and subsequent versions of DSpace, there is integration with Google Analytics via Google Analytics API. Google Analytics statistics are therefore accessible in DSpace – but you will need to configure DSpace to make this happen. See how to configure DSpace here.

To enable visits tracking, you will need to acquire a Google Analytics Key from Google – see Google Webmaster Tools. Add the key to the DSpace configuration in dspace.cfg under the setting jspui.google.analytics.key=UAXXXXXX-X (for JSPUI) or xmlui.google.analytics.key=UA-XXXXXX-X (for XMLUI) by replacing the UAXXXXXX-X with the key you will be given by Google Analytics. The statistics can be viewed on the Google Analytics web application at http://www.google.com/analytics/.

You can also consider the following for tracking content usage:
- The Content Usage Analysis module that gives you the tools to measure and report on the usage, content, growth, and therefore the value of your DSpace repository – a DSpace add-on from Atmire (not open source);
- MyDashfrom Harvard University usage reporting system – a stand-alone Open Source web application at https://github.com/oscharvard/mydash;
- Altmetrics such as those from Altmetric.com (not open source) or PlumX (not open source).

To get article level metrics, you could start with a free tool from Altmetric.com – Bookmarklet for Researchers. This tool will instantly give you article-level metrics for any recent paper (written and deposited after 2011).

To measure what type of attention research output has received, you could use Altmetric Institutional Repository badges, which are accessible with one click via the Altmetric donut

badges and give a record of all of the original shares and mentions of an individual piece of scholarly content. These badges are freely available for academic repositories and individual researchers. Embedding them is a straightforward process and badges are customizable.

How it works:
- Altmetric has to track your repository, items must include the appropriate metadata and the repository adds the badge code to its pages.
- Altmetric support a wide range of identifiers, including DOIs, PubMed Ids, ISBNs, Handles, arXiv Ids, ADS Ids, SSRN Ids, RePEC Ids, ClinicalTrials.gov records, URLs…so you can track attention gathered by items that have no DOI, if you like.

All the technical information you need is here.
To find out about the integration of altmetrics at DIGITAL.CSIC, the institutional repository of the Spanish National Research Council, see EIFL's webinar and PowerPoint presentation.

Find out more about Google Search and how to make your repository Google Scholar Search friendly.

☑ DOES YOUR REPOSITORY OFFER A GOOD USER EXPERIENCE FOR SMARTPHONES AND TABLETS?

Mirage 2 theme implements this capability by providing a distinct look for each of the three different categories of screen sizes: mobile phone, tablet and desktop.

Find out more in this webinar on enabling and customizing the Mirage 2 responsive XML User Interface for DSpace repositories.

☑ DOES YOUR REPOSITORY HAVE AN OPEN ACCESS REPOSITORY POLICY?

It is important for your own academic community, and the international community, to know what your repository is about. A comprehensive open access repository policy will help people to get to know your repository, understand how it works, and make it more visible.

Your institutional open access repository policy should define the overall vision of the repository, and include a collection policy, a submission policy, policy on the kinds of content to be included, a deposit licence and policy, a re-use licence, take-down policies and embargoes, a preservation policy, and rights, and responsibilities.

When you have a publicly stated open access repository policy for the permitted re-use of deposited items or for such things as submission of items, long-term preservation, etc, it simplifies matters for organizations wishing to provide search services, which in turn increases the visibility and impact of the repositories.

☑ RESEARCHER IDENTIFICATION - DO YOU ALLOW YOUR RESEARCHERS TO REGISTER TO DEPOSIT IN YOUR REPOSITORY WITH THEIR ORCID?

ORCID (Open Researcher and Contributor ID) is a non-proprietary alphanumeric code to uniquely identify scientific and other academic authors. The ORCID integration adds ORCID compatibility to the existing solutions for Authority control in DSpace.

More information: http://wiki.lib.sun.ac.za/index.php/SUNScholar/Researcher_Identification; http://wiki.lib.sun.ac.za/index.php/SUNScholar/Researcher_Identification/5.X/ORCID and https://wiki.duraspace.org/display/DSDOC6x/ORCID+Integration.

☑ NEW PLUGIN: OPEN PEER REVIEW MODULE FOR REPOSITORIES

With the support of OpenAIRE, Open Scholar coordinated a consortium of five partners to develop the first Open Peer Review Module (OPRM) for open access repositories in 2015-2016.

The OPRM has been developed as a plugin for repositories using the DSpace software package. Read more about the project here and download the software code here.

The OPRM envisions the gradual conversion of existing open access repositories into fully-functional evaluation platforms that will provide the capacity needed to bring back research quality control to the community of scholars and help bridge the gap between academic institutions and publishers.

See this webinar in which Isabel Bernal (DIGITAL.CSIC manager) presents the experience of integrating Open Peer Review Module DIGITAL.CSIC.

Further resources on how to set up, customize and manage your DSpace repository

Take a look at the expert tips for setting up and managing a DSpace repository, and resources on the following topics: DSpace installation, post-installation tasks, DSpace system administration and customizing DSpace.

Next generation repositories

On November 28, 2017 the Confederation of Open Access Repositories (COAR) released a report on next generation repositories. EIFL is one of the founding members of COAR.

Titled Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group, the report describes 11 new behaviours, as well as the technologies, standards and protocols that will facilitate the development of new services on top of the collective network, including social networking, peer review, notifications, and usage assessment:
1. Exposing Identifiers
2. Declaring Licenses at a Resource Level
3. Discovery through Navigation
4. Interact with Resources (Annotation, Commentary and Review)
5. Batch Discovery
6. Resource Transfer
7. Collecting and Exposing Activity Metadata
8. Identification of Users
9. Authentication of Users
10. Exposing Standardized Usage Metrics
11. Preservation of Resources

Read more about COAR's work on Next Generation Repositories.