



# HOW TO MAKE YOUR OA REPOSITORY WORK REALLY WELL

---

Checklist to help you get the best out of your DSpace repository

Iryna Kuchma, Open Access Programme Manager, EIFL

July 2019



Creative Commons Attribution 4.0 International Licence

# ***Table of Contents***

<b><i>Table of Contents</i></b> .....	2
Introduction .....	3
Which version of the DSpace software do you use? .....	3
Have you enabled the handle service? .....	3
Do you have backup/restore procedures and policies for disaster recovery? .....	4
Have you enabled the OAI-PMH server? .....	4
Do you follow standardized approaches to metadata?.....	5
How to improve discoverability through search engines .....	5
How to ensure that your repository is discoverable.....	6
Do you track and analyze statistics and web traffic? .....	7
Does your repository offer a good user experience for smartphones and tablets? .....	8
Does your repository have a policy? .....	9
Making your repository open: how to license repositories? .....	9
Researcher identification .....	10
Golden rules for repository managers from BASE .....	10
Open Peer Review Module for repositories .....	12
CORE Recommender.....	12
Next Generation repositories .....	13
Further resources on how to set up and manage your dspace repository .....	14

## INTRODUCTION

One of the EIFL Open Access Programme goals is to enhance open access (OA) 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 OA repositories that are built with DSpace software. This is a living document that we update periodically. It started with tips 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 OA repositories using DSpace free and open source software, which is the most commonly used repository software in EIFL partner countries. This is the fourth revision of the checklist.



### 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 6.3 that can be downloaded from [GitHub \(dspace-6.3 release\)](#). DSpace 6.3 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](#).

DSpace 6.3 upgrade instructions are available at [Upgrading DSpace](#). Nason Bimbe has prepared useful notes on how to upgrade from DSpace 5.x to 6.x with XMLUI, available [here](#).



### 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:

- Prefix registration: <http://www.handle.net/prefix.html>
- Payment: <http://www.handle.net/payment.html>
- Guides: <http://www.handle.net/documentation.html>  
[http://wiki.lib.sun.ac.za/index.php/SUNScholar/Handle\\_Server](http://wiki.lib.sun.ac.za/index.php/SUNScholar/Handle_Server)

- Video tutorial:  
<https://www.youtube.com/watch?v=sE5rBPRLS0Q&index=7&list=PLIOglq06moTsd63OaJn6zK2g1fV4sSwYv&t=0s>
- Documentation:  
<https://wiki.duraspace.org/display/DSDOC6x/Installing+DSpace#InstallingDSpace-TheHandleServer>



## 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 want to make sure it is backed up and monitored correctly.

Find out more about backup and recovery strategies and systems [here](#) and watch [this short tutorial](#).



## 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: If you need to change the default behaviour, you may need to alter the 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](http://wiki.lib.sun.ac.za/index.php/SUNScholar/Remote_Harvest).

Go to BASE [OAI-PMH validator OVAL](#) and validate your OAI-PMH to verify that everything was configured correctly.



## DO YOU FOLLOW STANDARDIZED APPROACHES TO METADATA?

Interoperability across repositories requires standardized approaches to metadata. Aim for as comprehensive metadata as possible. Try to include all descriptive information provided in the resource that you upload to your repository.

The [COAR Repository Toolkit](#) explains: “Ideally, repositories will expose their metadata using common schema and vocabularies so that the records can be standardized, and aggregated by repository networks. In turn, these networks can develop more useful services with the metadata, such as tracking OA, discovery of content, and analytics.”

“Currently most repositories expose their metadata through the Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH). This protocol allows the repository to use a variety of metadata profiles, in addition to the simple OAI-DC metadata format based on Dublin Core (DC). For generic data repositories, the DataCite metadata schema is the most widely used. Domain-based metadata schemas may also be used by repositories that specialize in collecting content from a specific discipline.”

“In addition, there are regional guidelines for repositories defined by certain repository networks, such as LA Referencia (Latin America) and OpenAIRE (Europe) require the adoption of certain specific metadata elements and vocabularies in order to provide services based on the metadata they aggregate.”

LIBSENSE initiative, coordinated by WACREN (West and Central African Research and Education Network), COAR and EIFL, outlined metadata guidelines for African repositories and recommended a use of controlled vocabularies - “lists of standardized terminology, words, or phrases, used for indexing or content analysis and information retrieval, usually in a defined information domain.” (CASRAI). The draft LIBSENSE document is available [here](#).



## HOW TO IMPROVE DISCOVERABILITY THROUGH SEARCH ENGINES

Search engines work by crawling web pages using their own web crawlers or search engine 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>.

There is more information in [this useful resource](#).

And check out the [recording](#) and [slides](#) from the COAR webinar with Kenning Arlitsch “Driving Traffic to Institutional Repositories: How Search Engine Optimization can increase the number of downloads from IR”.

For many researchers, Google Scholar is the first entry point for finding scientific articles. Check out the following resources:

- [Inclusion Guidelines for Webmasters](#)
- Recording and slides from DSpace and Google Scholar Webinars in which Google Scholar Monica Westin shows how Google Scholar indexing works and what is required to ensure the inclusion of repositories. The webinars cover the most common DSpace repository indexing issues, how they can be identified and resolved, and how to check the coverage of your repository in Google Scholar.
  - [DSpace and Google Scholar Webinar for Ghana](#)
  - [DSpace and Google Scholar Webinar for Uganda](#)



## HOW TO ENSURE THAT YOUR REPOSITORY IS DISCOVERABLE

How easy is it to find your repository? In order to ensure that content in your repository can be found, [register](#) it with [OpenDOAR](#), an authoritative listing of academic repositories.

Once registered with OpenDOAR, your repository can be picked up by other services such as [CORE](#), which aggregates OA content, benefiting those whose content is used and increasing its impact. As well as offering a web-based search engine, it also provides programmable access to metadata and full-text for text mining.

[Register](#) your repository with [BASE](#) – one of the world's most voluminous search engines especially for academic web resources.

Services and browser extensions like [Unpaywall](#) and [Open Access Button](#) – that locate freely available copies of research articles published in subscription-based journals and available in OA repositories – use CORE and BASE data.

If you want to make sure that content from your repository appears in Unpaywall, check whether it is listed among the data sources at <https://unpaywall.org/sources>. And if you'd like to add your repository to their list of sources, you can submit your repository for indexing via [this form](#).

Institutional repository managers can use Unpaywall data to find OA resources that faculty have posted online, but which they have not deposited in their institutional repository. These can be automatically ingested (included), significantly increasing institutional repository coverage without needing to convince faculty to deposit. For example, SwePub (national repository of Sweden) added 75,000 new OA records, increasing number of OA records by over 30%, while the smaller the Carleton University repository added 1,000 OA records, doubling their fulltext coverage. There are a few good ways to access Unpaywall data for this use case: [download the whole dataset](#), [check lists of records by DOI](#), or take advantage of Unpaywall's integration into [Dimensions](#), [Web of Science](#), and [Scopus](#).



## DO YOU TRACK AND ANALYZE STATISTICS AND WEB TRAFFIC?

Having good statistics about the 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, etc. (all the data is available in the SOLR indexes – these can be queried directly as well).

Use Google Analytics (registration required) or [Matomo](#) (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);
- MyDash from 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 slides and recording from the EIFL webinar](#).



## **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 A POLICY?

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

Your institutional OA 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 OA repository policy for the permitted re-use of deposited items, it simplifies matters for organizations wishing to provide content aggregation and indexing services, which in turn increases the visibility and impact of the repositories.

OpenDOAR provides [a tool to support the development of a repository policy](#).



## MAKING YOUR REPOSITORY OPEN: HOW TO LICENSE REPOSITORIES?

Apply the right licence to your repository. One of the best licences you can use for your repository is a [Creative Commons Attribution 4.0 International Licence](#) (CC BY 4.0), specifying that “unless otherwise noted, this repository is under a CC BY 4.0 licence”.

OpenAIRE [recommends](#) using a CC BY 4.0 licence as a repository licence for the following reasons:

- Creative Commons licences are internationally recognized, well-established, and both human-readable and machine-readable;
- CC BY 4.0 licences meet the definition of “open access” as defined in the Budapest, Bethesda, and Berlin declarations on open access;
- CC BY 4.0 is one of the most compatible licences for interoperability purposes.



## RESEARCHER IDENTIFICATION

[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); [http://wiki.lib.sun.ac.za/index.php/SUNScholar/Researcher\\_Identification/5.X/ORCID](http://wiki.lib.sun.ac.za/index.php/SUNScholar/Researcher_Identification/5.X/ORCID) and <https://wiki.duraspace.org/display/DSDOC6x/ORCID+Integration>.

DSpace 6.3 has been released with the following fix for ORCID Integration: Update DSpace ORCID Integration to use ORCID API v2 (instead of now obsolete ORCID v1): [DS-3447](#)



## GOLDEN RULES FOR REPOSITORY MANAGERS FROM BASE

BASE has compiled a useful list of [golden rules](#) that will help you optimize your repository discoverability.

- **OAI interface working**

Your OAI interface is stable and responds to requests. *ListRecords* delivers results without timeout or other issues, e.g. an XML error.

Otherwise, it is not possible to index your repository.

- **Marking modified / deleted records**

Any subsequent change to a record must be marked in your OAI interface by updating the document's datestamp. If a document is deleted from your repository, the record in the OAI interface must be marked as "deleted". Under no circumstances may the record be completely deleted from the OAI interface.

- **Comprehensive metadata**

Each item exposed via your OAI interface provides metadata as comprehensively as possible (title, author, abstract, publication date) using the [info-eu-repo vocabulary](#).

- **Working / persistent identifier (URL, DOI, Handle, URN)**

Each record contains a working URL in <dc:identifier> (starting with http or https). This directs either to the frontdoor of the document or directly to the full text (PDF). If the full text

is not provided in a common file format (HTML, PDF) the identifier should direct to the frontdoor and not directly to the full text.

Preferably provide persistent identifiers (DOI, Handle, URN) which will still be working after relocation of the server and thereupon change of the URL. Make sure that DOIs etc. are registered with an appropriate registration agency and that they are resolving. DSpace installations in particular will need to reconfigure the 'handle', or it will refer to a dummy URL (123456789) which generates an error message.

- **Providing access information (Open Access)**

Access information of the full text document should be provided in the field <dc:rights> of each item according to the [info-eu-repo vocabulary](#). Alternative: OA documents are provided in an extra set (OA set). The name of this set is listed in each metadata record in the field *setSpec*.

- **Providing information concerning re-use / licence (Creative Commons)**

Authors can publish their work under a Creative Commons licence in your repository. You expose the chosen licence in your OAI interface within an additional <dc:rights> field, e.g. <dc:rights><https://creativecommons.org/licenses/by/4.0/></dc:rights>.

- **Add an ORCID iD to author names (and other person identifiers if possible).**

Promote the adoption of [ORCID iDs](#) (and other person identifiers if possible) to uniquely identify authors (even in case of name ambiguity). Encourage your researchers to register with ORCID in order to obtain an ORCID iD and add the ORCID iDs to the metadata. Some metadata formats delivered via OAI-PMH display ORCID iDs in a separate tag and therefore can be indexed separately. So far, BASE indexes exclusively OAI-DC (Dublin Core). Here, ORCID iDs should be provided directly as part of the author's name (e.g. <dc:author>Summan, Friedrich (ORCID-ID 0000-0002-6297-3348)</dc:author>).

If an ORCID iD exists, authors can also be found using the ORCID iD.

- **Character encoding**

All content exposed via your OAI interface (title, creator, abstracts) is encoded in [UTF-8](#).

- **Publication date**

The publication year or publication date is provided in the field <dc:date> in [ISO 8601](#) (YYYY-MM-DD, e.g. 2016-04-01 for the 1st of April, 2016) according to the [Gregorian \(western\) calendar](#). The field <dc:date> should only be used once.

- **Document language**

The language of a document is provided in the field <dc:language> in [ISO 639](#) (2 or 3 letter code, e.g. *en* or *eng* for English).

- **Source / Suggested citation**

The source or suggested citation of an item (e.g. journal's name, volume and issue of an journal article) is provided in <dc:source>. These details allow a better retrieval of your documents.

- **Items per page**

Every *ListRecords* includes 50-1000 items at most. The *resumptionToken* is working and is delivering the next 50-1000 items.

- **Contact person**

The *identify* request of your OAI interface includes the field *adminEmail*, which contains the active e-mail address of a technical admin. The homepage of your repository includes the e-mail address of the content provider.

## 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 OA 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 OA 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](#).



## CORE RECOMMENDER

CORE offers [a repository plugin](#), which provides recommendations on related publications to users. The recommender is a plugin that can be installed in repositories to suggest similar full-text items available in your or other OA repositories. Its purpose is to support users in finding articles relevant to what they read. CORE has validated that all recommended articles are free to read and can be accessed without a paywall. Its state-of-the-art recommendation algorithm

makes use of article full texts, not just metadata, in addition to a wide range of other features. Read more on how the algorithm works in the [CORE research section](#).

For those with access to the [CORE Repository Dashboard](#) – the Recommender installation guidelines and an installation key can be accessed via the Dashboard. Log into the Dashboard and then choose the tab *Get the recommender*.

Otherwise, visit CORE [registration page](#) to gain access to the Recommender.

Find out more about the CORE Recommender in this [CORE blog post](#).

## NEXT GENERATION REPOSITORIES



On 28 November 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 Licences 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

COAR's Next Generation Repositories (NGR) project positions repositories as the foundation for a distributed, globally networked infrastructure for scholarly communication. NGR technical vision is to make the resource, rather than the repository, the focus of services and infrastructure. A Next Generation repository manages and provides access to a wide diversity of resources; is resource-centric; is a networked repository; is machine-friendly; and is active.

Read more about [COAR's work on Next Generation Repositories](#).

Browse through the behaviours and technologies on the [Next Generation Repositories website](#).

Two supporting technologies are worth highlighting - ResourceSync and Signposting.

[ResourceSync](#) is a synchronization framework for the web consisting of various capabilities that allow third-party systems to remain synchronized with a server's evolving resources. As a successor of the OAI-PMH protocol it is faster, more reliable, scalable and allows real-time notification (and recovering of missed messages).

4Science has released a ResourceSync implementation for DSpace 5, 6 and (future) 7 with the support of OpenAIRE.

[Signposting](#) is an approach to make the scholarly web friendlier to machines exposing relations as Typed Links in HTTP Link headers, fully aligned with hypermedia (REST, HATEOAS) lines of thinking regarding web interoperability.

Signposting is now implemented in DSpace-CRIS and OJS. DSpace 7 plans to provide Signposting support.

## **FURTHER RESOURCES ON HOW TO SET UP 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](#).