



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

December 2021



Creative Commons Attribution 4.0 International Licence

Table of Contents

Table of Contents.....	2
Introduction	3
Which version of the DSpace software do you use?.....	3
Have you enabled the handle service?.....	4
Do you have backup/restore procedures and policies for disaster recovery?.....	4
Have you enabled the OAI-PMH server to ensure interoperability?	4
Do you follow standardized approaches to metadata?.....	5
Have you ensured search engine optimization?	6
How discoverable is your repository?	7
Do you track and analyze statistics and web traffic?.....	8
Does your repository offer a good user experience for smartphones and tablets?.....	10
Does your repository have a policy?.....	10
have you applied the right licenCe to your repository?.....	11
Have you integrated ORCID in your repository?	11
Appendix 1 Golden rules for repository managers from BASE.....	12

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 fifth 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 7.1, providing new features, security and bug fixes to the 7.x platform.

DSpace 7 is the largest release in the history of DSpace software. DSpace 7 introduces a brand new (Angular-base) User Interface (UI), which brings together the best of both the JSPUI and XMLUI worlds – two different UIs that prior versions of DSpace (v6.x and below) used and which are no longer supported in v7. This new UI is also backed by a brand new REST API, which opens all data and features to the web, allowing DSpace to integrate or interact with external systems/services. DSpace 7.1 [Latest Release page](#) provides download and installation links and new major feature highlights. Many more new and redesigned features are listed in these [Release Notes](#). Much more information is also available in the [DSpace 7 Documentation](#). Full installation instructions are available at [Installing DSpace](#). Upgrading instructions are available at [Upgrading DSpace](#).

The previous version of software – DSpace 6.3 can be downloaded from [GitHub \(dspace-6.3 release\)](#) and 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](#).

Please note that DSpace 4x isn't any longer supported and you should upgrade your software as soon as possible.



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
- Video tutorial: <https://youtu.be/sE5rBPRLS0Q>
- 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 TO ENSURE INTEROPERABILITY?

“The real value of repositories lies in the potential to interconnect them to create a network of repositories, a network that can provide unified access to research outputs and be (re-) used by machines and researchers. However, in order to achieve this potential, we need interoperability, ...the ability for systems to communicate with each other and pass information back and forth in a usable format... Within repositories, interoperability occurs in multiple ways. At the system level, interoperability occurs when repositories are set up in such a way that they allow for data or digital objects to be passed into or out of repositories via external systems. One example, the Protocol for Metadata Harvesting (OAI-PMH), specifies certain criteria that must be met in order to allow external, third-party systems to access and harvest metadata from repositories.” ([The Case for Interoperability for Open Access Repositories, COAR](#))

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

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](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.

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 also 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 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](#). See also Appendix 1: A set of 'golden rules' from BASE.

The use of controlled vocabularies for bibliographic metadata "ensures that everyone is using the same word to mean the same thing". The [COAR Controlled Vocabularies](#) offer the [Resource Type vocabulary](#), the [Access Rights vocabulary](#) and [the Version Type vocabulary](#). They could be implemented in repositories operating on DSpace by customizing input forms or using the Dspace software functionality, see a step by step guide [here](#). And if you want to translate the vocabularies in your language, please [get in touch!](#)



HAVE YOU ENSURED SEARCH ENGINE OPTIMIZATION?

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 Monica Westin of Google Scholar 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 DISCOVERABLE IS YOUR REPOSITORY?

How easy is it to find your repository?

OpenDOAR

In order to ensure that content in your repository can be found, [register](#) it with [OpenDOAR](#), the global Directory of Open Access Repositories.

CORE

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, CORE also provides programmable access to metadata and full-text for text mining.

You could also directly [register](#) your repository with CORE.

CORE offers [a repository plugin](#), which provides recommendations on related publications to users. It can be installed in repositories to suggest similar full-text items available in your or other OA repositories. CORE validates that all recommended articles are free to read and can be accessed without a paywall. 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 the CORE [registration page](#) to gain access to the Recommender. Find out more about the CORE Recommender in this [CORE blog post](#).

BASE

[Register](#) your repository with [BASE](#) – one of the world's most voluminous search engines especially for academic web resources. (See also Appendix 1: A set of 'golden rules' from BASE that will help you optimize your repository discoverability.)

Unpaywall

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 the number of OA records by over 30%, while the smaller Carleton University repository added 1,000 OA records, doubling their full text coverage. There are a few good ways to access Unpaywall data for this use case: [download the whole dataset](#) or [check lists of records by DOI](#).

OpenAIRE

With OpenAIRE registration you could reach a wider audience around the world and be part of an interlinked repository network. OpenAIRE's Notification Broker Service helps to enrich your metadata. Read more in these two guides: '[Provide – How to validate and register your data source](#)' and '[Provide – How to enrich research artefacts](#)', and sign in for the Provide Dashboard [here](#).

Wikipedia

Do you know how many Wikipedia pages are linking to your repository content? To check, go to the External links search page <https://en.wikipedia.org/wiki/Special:LinkSearch> and enter the url of your repository to identify which Wikipedia pages link to your repository content. This page exclusively searches the English Wikipedia and to search the instances of Wikipedia in other languages, just copy the path /w/index.php?target=&title=Special%3ALinkSearch or /wiki/Special:LinkSearch in the Wikipedia instance of choice. Here is a French example: https://fr.wikipedia.org/w/index.php?title=Sp%C3%A9cial:Recherche_de_lien&target=. Wikipedia increases the visibility of your content and could be one of the societal impact indicators. Check out [this Atmire article](#) and a tutorial video with more details and tips.



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, if available); 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=UAXXXXXXX-X (for JSPUI) or xmlui.google.analytics.key=UA-XXXXXX-X (for XMLUI) by replacing the UAXXXXXXX-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.

See more details [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 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 and aims of the repository, and include a content policy that describes accepted types of content; a submission policy; access, use and reuse policy for content and metadata, including a deposit and re-use licence; take-down policy, and a preservation policy.

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.

Use RePol - Repository Policy Generator at <https://repol.ni4os.eu> to draft your policy (more information about this tool is in [this EIFL's webinar](#)).

If you have doubts about publisher copyright policies and permissions for uploading content in repositories, check [Sherpa Romeo](#) online resource that aggregates and analyses publisher OA policies from around the world and provides summaries of publisher copyright and OA archiving policies on a journal-by-journal basis. You could also use [Shareyourpaper.org](#) to automate repository deposit workflows for journal articles (more details are in [this webinar](#)).



HAVE YOU APPLIED THE RIGHT LICENCE TO YOUR REPOSITORY?

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.

[COAR Community Framework for Good Practices in Repositories](#) recommends that metadata in a repository are available under a Creative Commons Public Domain License and are available for download in a standard bibliographic format.



HAVE YOU INTEGRATED ORCID IN YOUR REPOSITORY?

[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 and integration tips:

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

Read more about DSpace 7 ORCID integration [here](#). And find practical advice, ideas and tips for repository managers on ORCID and repository integrations in [this webinar](#).

APPENDIX 1

GOLDEN RULES FOR REPOSITORY MANAGERS FROM BASE

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

- **Functioning OAI interface**

Your OAI interface is freely accessible, stable and constantly responding. The request for ListRecords in oai_dc format returns results without a timeout or output error. You should check the functionality of your OAI interface at regular intervals, e.g. using a browser. If your OAI interface does not function correctly, it is not possible to index your repository.

- **Contact Persons**

In the Identify data of your OAI interface, an e-mail address is specified in the adminEmail field, which can be used to contact the technical operator of the OAI interface. An e-mail address is available on the homepage of your repository, which guarantees direct contact.

- **Character encoding**

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

- **URL of the publication <dc:identifier>**

Each record contains a working URL in the field <dc:identifier> (starting with http:, https: or doi). Provide persistent identifiers (DOI, handle, URN) that will continue to function even if the server is relocated and the URL is changed. Especially for DSpace installations the handle has to be configured, otherwise it will lead to a "dummy-URL" (handle.net/123456789), which generates an error message.

Code Examples

```
<dc:identifier>http://hdl.handle.net/10760/12746</dc:identifier>
<dc:identifier>https://doi.org/10.1108/07378830610715473</dc:identifier>
<dc:identifier>doi:10.1108/07378830610715473</dc:identifier>
```

- **Title <dc:title>**

Enter the title in the <dc:title> field as in the original. If the publication has several titles (e.g. in different languages), repeat the field.

Code Example

```
<dc:title>Advanced calculus: student handbook</dc:title>
```

- **Publication date <dc:date>**

Each record should contain in the <dc:date> field the publication year or date of the document in [ISO 8601](#) format (according to the [Gregorian calendar](#)). The field <dc:date> should only be filled once. If there is no concrete publication date, estimate. Inaccurate data such as 17th century should be given as 1650.

Code Examples

```
<dc:date>2000-12-25</dc:date>
<dc:date>1978-02</dc:date>
<dc:date>1650</dc:date>
```

- **Language of the document <dc:language>**

You provide information on the language of a document according to [ISO 639](#) (2- or 3-letter code) in the <dc:language> field.

Code Examples

```
<dc:language>eng</dc:language>
<dc:language>en</dc:language>
```

Access and re-use rights <dc:rights>

- **Access rights (Access status)**

The <dc:rights> field contains access information to the full text. Use [the COAR Access Rights vocabulary](#). The alternative is: Open Access documents are available in their own OAI set. The name of this set is contained in the setSpec field for each record. Name the set as uniquely as possible, e.g. open access.

Code Examples

```
<dc:rights>closed access</dc:rights>
<dc:rights>http://purl.org/coar/access\_right/c\_abf2</dc:rights>
```

- **Subsequent use rights (licenses)**

Use licenses that are as widely used as possible, such as Creative Commons licenses. Enter the corresponding license in your OAI interface in another <dc:rights> field.

Code Example

```
<dc:rights>https://creativecommons.org/licenses/by/4.0/</dc:rights>
```

- **Reference / Citation <dc:source>**

Information about the source or the citation (e.g. for articles the name, volume, issue of the journal) can be found in <dc:source>.

Code Examples

```
<dc:source>Ecology Letters (1461023X) vol.4 (2001)</dc:source>
<dc:source>ISSN: 0928-0987</dc:source>
<dc:source>Pieper D, Summann F.: Bielefeld Academic Search Engine (BASE). An end-user oriented institutional repository search service. Library Hi Tech. 2006; 24(4):614–619. ISSN 0737-8831.</dc:source>
```