Redistributable Repositories

This page contains a list of repositories with contents which you can download and put into your own repository. Make sure to check the terms and conditions of each repository. Some may allow redistribution of certain items, but not others.

Testing data

If all you need is data for internal testing (not public), you don't need to restrict yourself to one of the repositories listed here. Most DSpace repositories offer the OAI-PMH interface, so you can harvest their metadata and in some cases even their bitstreams (only bitstreams with anonymous read access; only DSpace repositories using XMLUI).

Remember that to display the data publicly on the internet, you need a license (permission from the copyright holder) for redistribution. See below for a List of redistributable repositories.

Locating the OAI-PMH interface

The OAI interface is usually accessible by adding "/oai/request" to the repository domain name or URL. Example of OAI base URL:

```
http://example.com/oai/request
```

To verify whether the OAI repository really is accessible, you should add "?verb=Identify" to the base URL to get an XML document with repository description. Example:

```
http://example.com/oai/request?verb=Identify
```

If it returns the XML document, you have found an OAI interface, proceed with next chapter How to harvest. If it returns a page not found (404 HTTP response code) or an error page, there is no OAI interface at this address (either the repository doesn't provide an OAI-PMH interface or it's not accessible at this URL or it's not publicly accessible).

How to harvest

Official documentation:

- OAI-PMH / OAI-ORE Harvester (reference)
- Harvesting Items from XMLUI via OAI-ORE or OAI-PMH (howto)

List of redistributable repositories

More information can be found in the http://openbiblio.net/

Repo name	license / Terms of use	OAI interface / data files	number of items
Pub Med Central's Open Access Subset	CC licenses	OAI-PMH & FTP (for files)	over 1 million
OKR Worldbank	Terms of use	OAI-PMH	?
ShareGeo Dev	most should be redistributable (TODO: verify)	OAI-PMH	?
British Library	CCo Public Domain Dedication License	CKAN Package	3 Mio (2010-11)
Swedish National Bibliography	CCo	OAI-PMH feed	2.4 Mio
Europeana Linked Open Data	Data Exchange Agreement (DEA) CCO	Datasets and own API	20 Mio
Spanish National Library		Datahub	
German National Bibliography	CC0	OAI SRU Datahub	11 Mio
Research Data Curation Bibliography	CC-BY-NC 3.0	N/A	650

(When we have a larger list and we'll have checked the license terms, we can harvest the content and provide convenient dumps in form of SQL files and AIP or Simple Archive Format files here.)

TODO: http://blog.sense.io/23-resources-for-finding-open-data/