OAI-PMH Metadata Delivery for Catalogs and Institutional Repositories

OAI-PMH Metadata Delivery:

Institutional Repository databases are created by harvesting the customer’s metadata using a protocol called OAI-PMH. The Institutional Repository Questionnaire asks the customer to provide their institution’s OAI-PMH URL that retrieves the records from their repository or local archives.

Most institutional or digital repository software vendors support Dublin Core metadata export and therefore can utilize OAI-PMH harvest. [Note: juni2 metadata can also be harvested via OAI-PMH]

Below are links to the most common vendor’s information regarding OAI-PMH support and/or harvest instructions.

• **Aleph** - There are no specific instructions for OAI harvest on their web site.
• **DSpace** - [https://wiki.duraspace.org/display/DSDOC18/OAI](https://wiki.duraspace.org/display/DSDOC18/OAI)
• **Digital Commons** – There are no specific instructions for OAI harvest on their web site. This is in their FAQ:
  ◦ Does Digital Commons support the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)? Yes. Digital Commons supports OAI-PMH version 2.0.
  ◦ Is Digital Commons an OAI Data Provider, or Service Provider? Digital Commons is an OAI Data Provider, but not a Service Provider. This means that Digital Commons sites support the OAI Protocol for Metadata Harvesting (OAI-PMH) as a means of exposing metadata, but the sites do not harvest OAI data from other sites.
• **ePrints** - [http://wiki.eprints.org/w/OAI](http://wiki.eprints.org/w/OAI)

Common technical issues with OAI-PMH harvests:

**Server not found:** The customer’s OAI server must be available at all times to ensure that the EBSCO harvester can connect during our business hours.

**Double encoding and/or html encoding:** Encoding of the XML must use the UTF-8 representation of Unicode. Character references, rather than entity references, must be used. Character references allow XML responses to be treated as stand-alone documents that can be manipulated without dependency on entity declarations external to the document.

**Incomplete harvest:** Flow control, through the use of incomplete list responses and resumptionToken elements, is a powerful mechanism that allows a repository to throttle incoming requests from harvesters. The size of a list response
can result in an incomplete-list response. See the OAI Implementation Guidelines for more information:
http://www.openarchives.org/OAI/2.0/guidelines-repository.htm.

**No sets:** If there are no sets configured in the implementation, this must be indicated in the customer questionnaire. The harvester has to be configured with to harvest data successfully. If sets are configured, a sets.xml file must be present.

**Incorrect URLs:** Please test the harvest URL to ensure that it is correct. This OAI Repository Explorer provides a test interface: http://re.cs.uct.ac.za/. EBSCO’s harvester does not support the “&rt=” parameter.

**Related FAQ:**

EBSCO Discovery Service Institutional Repository Database Questionnaire