

Selective Harvesting: Creating and Ingesting Custom OAI-PMH Sets

Author/Contributor:

Sidhunata, Harry R.; Croucher, Joanne L.; Frances, Maude

Event details:

4th eResearch Australasia Conference
Gold Coast, Australia

Publication Date:

2010

DOI:

<https://doi.org/10.26190/unsworks/1133>

License:

<https://creativecommons.org/licenses/by-nc-nd/3.0/au/>

Link to license to see what you are allowed to do with this resource.

Downloaded from <http://hdl.handle.net/1959.4/45545> in <https://unsworks.unsw.edu.au> on 2022-07-05

Selective Harvesting: Creating and Ingesting Custom OAI-PMH Sets

Harry R. Sidhunata, Joanne L. Croucher, Maude Frances; University Library, University of New South Wales, Sydney, Australia

DEFINITION

The Selective Harvester provides a flexible mechanism to aggregate, filter and re-use metadata records from multiple open access repositories. It integrates an existing OAI tool (jOAI)¹ with an application that filters and ingests selected records into a Fedora Repository².

USE CASE

A key use case for selective harvesting is to help populate subject-based repositories. By aggregating resources on a particular topic, these repositories act as search and discovery portals.

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)³ has been widely adopted to disseminate metadata records and support interoperability between repositories. OAI-PMH supports selective harvesting by date or by set. However establishing sets can be a burden for data providers, especially as they may be harvested by many different service providers. A flexible, independent mechanism to filter, review and re-use harvested records will support subject-based repositories without requiring data providers to set up multiple, complex OAI sets.

DESIGN

The Selective Harvester enables the creation of custom OAI-PMH sets and manages the import of selected records into a Fedora repository. The model consists of two separate Java-based open source applications, each of which may be deployed and configured independently.

1) jOAI

This OAI harvester and provider was developed by Digital Learning Sciences (DLS) at the University Corporation for Atmospheric Research. Highly configurable, it enables the harvesting of multiple open access repositories as well as the specification of OAI-PMH sets.

2) Filter & Ingest Modules

Two independent modules have been developed at the University Library, University of New South Wales (UNSW). The Filter Module can be used to filter and review harvested records. The Ingest Module has the facility to transform the harvested records prior to import into a Fedora repository. If required, the Ingest Module can also operate independently of both jOAI and the Filter Module.

FEATURES

- Open source Java-based applications
- Flexible configuration
- Harvest one or more open access OAI-compliant repositories
- Support for different XML metadata formats
- Define advanced search criteria to filter harvested resources
- Preview filtered resources and option to manually review resources
- Scheduling facility (components can be set up to operate automatically at set intervals)
- Option to upload and use XSL files to transform the harvested metadata
- Ingest harvested and filtered resources to a Fedora-based repository

IMPLEMENTATION

The Selective Harvester has been implemented by the NCHSR Clearinghouse⁴, a subject-based repository developed jointly by UNSW Library and researchers at the National Centre in HIV Social Research (NCHSR). The NCHSR Clearinghouse aggregates local resources with selected records from an institutional repository.

DOWNLOAD

The Selective Harvester Filter and Ingest Modules are written in Java, and available as Open Source software via SourceForge: <http://sourceforge.net/projects/selectharvest/>

References

1. jOAI Overview: The Java-based Open Archives Initiative Data Provider & Harvester, Available from: <http://www.dlese.org/oai/>
2. Fedora Commons Repository Software, Available from: <http://fedora-commons.org/>
3. Open Archives Initiative: Protocol for Metadata Harvesting (OAI-PMH), Available from: <http://www.openarchives.org/pmh/>
4. NCHSR Clearinghouse, Available from: <http://ssrm.nchsr.arts.unsw.edu.au>

More Information

Selective Harvesting: Selecting and Ingesting OAI-PMH Sets, Available from: <http://handle.unsw.edu.au/1959.4/45545>

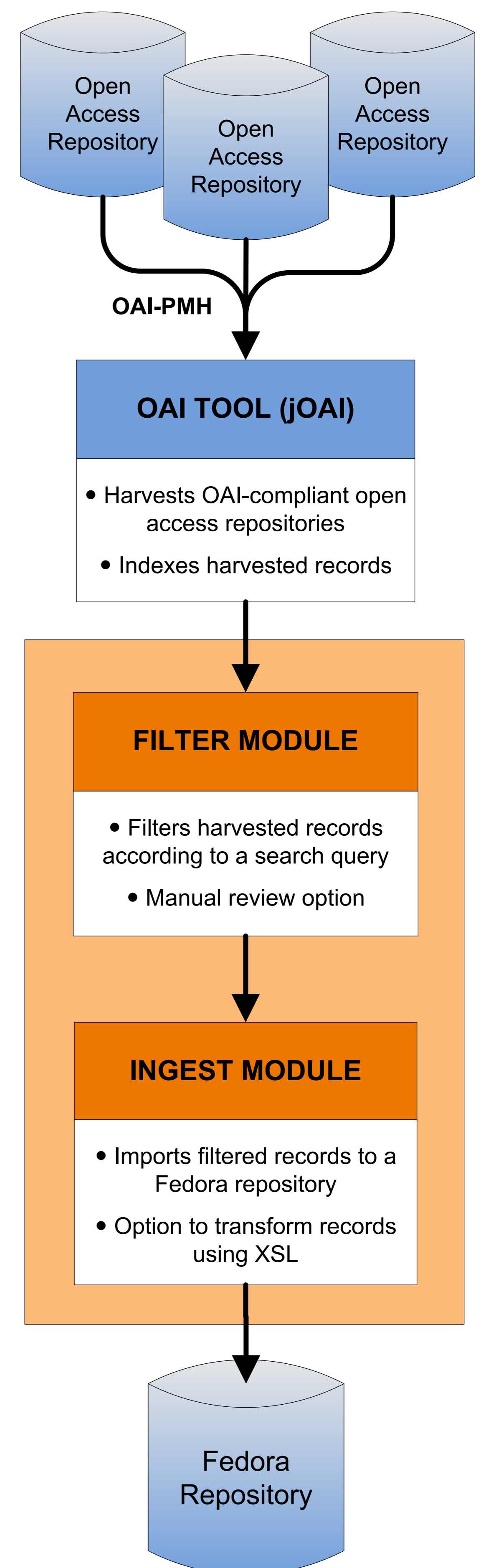


Figure 1: Selective Harvester Model

Acknowledgements

The Selective Harvester was developed at the University Library, University of New South Wales. This product includes software developed by Digital Learning Sciences (DLS) (<http://www.dlsciences.org/>) at the University Corporation for Atmospheric Research (<http://www.ucar.edu/>).

Contact

Harry Sidhunata
Technical Support Officer
Library Repository Services, University Library, UNSW
Email: h.sidhunata@unsw.edu.au

Jo Croucher
Support Officer
Library Repository Services, University Library, UNSW
Email: j.croucher@unsw.edu.au