

Harvesting to Aggregators in Islandora

Moderator:
Erin Tripp, discoverygarden
erin@discoverygarden.ca
[@eeohalloran](https://twitter.com/eeohalloran)

Panelist:
Giancarlo Birello, CNR-IRCrES
giancarlo.birello@ircres.cnr.it

Panelist:
Nick Ruest, York University
ruestn@yorku.ca
[@ruebot](https://twitter.com/ruebot)

Panelist:
Jordan Dukart, discoverygarden
jordan@discoverygarden.ca
[@discgarden](https://twitter.com/discgarden)

Agenda

Harvesting Introduction

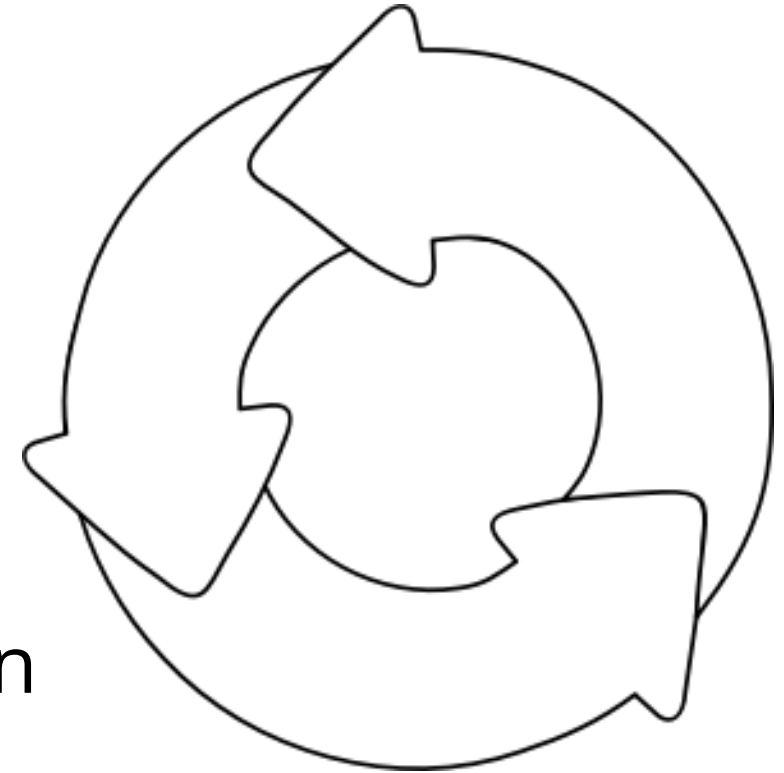
Panelist Introductions

Question Period

Introduction - Harvesting

Harvesting Overview

- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
- Low-barrier mechanism for repository interoperability
- OAI-PMH is a set of six requests (aka verbs or services) that are invoked within HTTP



<https://www.openarchives.org/pmh/>

Harvesting Roles

- *Data Providers* are repositories that expose structured metadata via OAI-PMH
- *Service Providers* then make OAI-PMH service requests to harvest that metadata

Harvesting Impacts

February 2015 LITA panelists say Top Technology Trends include enhancing **discoverability** [\(Enis, 2015\)](#)

Making content accessible **where the search originates** (e.g. Google, Google Scholar, WorldCat, DPLA, Europeana) creates value for digital libraries and users [\(Enis, 2015\)](#)

Repositories contributing to aggregators can experience **increased site visits from 55-109 per cent** [\(DPLA, n.d.\)](#)

Being a Data Provider

Islandora is OAI-PMH compliant

Users of Islandora OAI module are being harvested by

- WorldCat
- Europeana
- DPLA (and other national aggregators - Trove, Cultura Italia, etc.)

Module written so that portions can be overwritten for custom implementations

Introductions - Our Panelists

Nick Ruest

Nick Ruest is the Digital Assets Librarian at York University. He oversees the development of data curation, asset management and preservation initiatives, along with creating and implementing systems that support the capture, description, delivery, and preservation of digital objects having significant content of enduring values.

He is also active in the Islandora community, serving as the Release Manager, Project Director for the Islandora and Fedora 4 integration project, member of the Islandora Foundation's Roadmap Committee, and contributes code to the project.

In the past he has served as the moderator for the OCUL Digital Curation Community, the President of the Ontario Library and Technology Association, and McMaster University Academic Librarians' Association.

York University

York University Digital Library uses the Islandora OAI module primarily for harvesting by the OCLC Digital Collections Gateway.



Nick worked with Jordan on providing generalized modifications to the module that allows for links to the object thumbnails, and links back to the object in Worldcat.

Check out the harvested data at:

http://www.worldcat.org/search?q=on%3ADGCNT+https%3A%2F%2Fdigital.library.yorku.ca%2Foai2+YOU&qt=results_page

A screenshot of a WorldCat search results page. It shows two entries for digital collections. The top entry is titled "Steacie Library Dungeon Hackfest 2014" by York University Libraries (Creator). It is labeled as "Downloadable archival material", in English, published by York University Libraries on 2015-01-13, and is in the WorldCat database. The bottom entry is titled "Steacie Science and Engineering Library » The Steacie Library Dungeon Hackfest 2013" by York University Libraries (Creator). It is also labeled as "Downloadable archival material", in English, published by York University Libraries on 2015-01-13, and is in the WorldCat database. Each entry includes a small thumbnail image of the collection's cover.

Giancarlo Birello

Giancarlo Birello is the technical manager of the IT Office at Italy's Research Institute on Sustainable Economic Growth of National Research Council (aka CNR IRCrES) with the main role of network and system manager for CNR Piedmont ITC infrastructure.

He managed the development of the Digibess Islandora repository, a project that handles digitized objects from 18 Economic and Social Sciences libraries in Italy. It was harvested by the Italian aggregator Cultura Italia and the European aggregator, Europeana. He's also involved in the new V2P2 Islandora repository focused on improving access to agricultural data internationally.

Giancarlo's directly managed harvesting initiatives by national and international aggregators regarding protocol and metadata issues.

CNR IRCrES

Islandora OAI module will be used by the V2P2 project. It's in the prototype stage currently. Previous experience with the Digibess OAI initiative allowed examination of using the protocol, such as:

- How SETS are used to satisfy aggregator requirements, how SETS can be associated to collections and SETS naming policy
- What type of metadata is needed and how basic metadata must be managed for different harvesters
- The use of intermediate aggregator (e.g. Culturaitalia to Europeana)
- Solving metadata issues is paramount harvesting to aggregators and creating an open and useful repository

See the content on Europeana at: http://www.europeana.eu/portal/search.html?qt=false&rows=24&qf=DATA_PROVIDER%3ADigibess



Matches for:

By data provider: Digibess 

Jordan Dukart

A discoverygarden developer since 2012, and a development team lead since 2014, Jordan's the current maintainer of the Islandora OAI module. He admittedly knows far too much about the OAI2 spec.

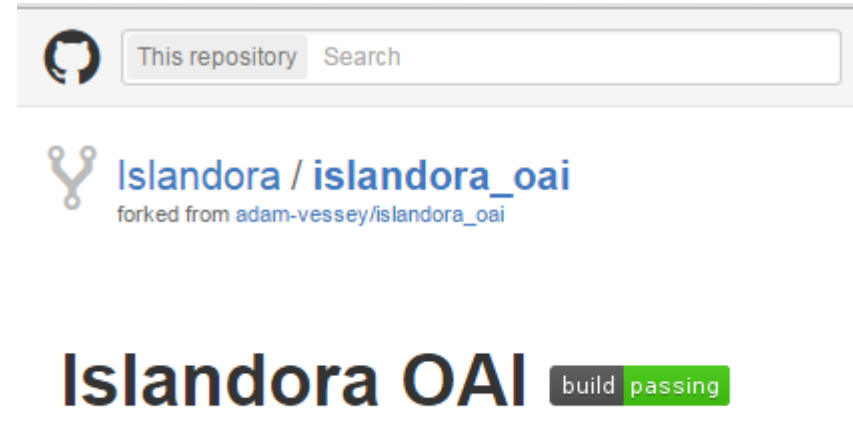
The Islandora OAI module provides an implementation the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) for an Islandora (Fedora) repository with a Solr index. By implementing the islandora_oai module, you can expose content -- a repository's metadata -- as an OAI-PMH repository, allowing it to be accessible by OAI harvesters.

Islandora OAI Maintainer

Islandora OAI is one of the most widely touched modules within Islandora. It has eighteen committers (code contributors).

Islandora OAI was the first project Jordan worked on at discoverygarden. Since day one, becoming and staying OAI compliant as well as meeting WorldCat's entrance requirements has been a central goal.

Iteratively, the underlying module has been re-worked such that each portion of the vocabulary can be overwritten as needed or changed to support non-standard data models as needed.



The screenshot shows the GitHub repository page for 'Islandora / islandora_oai'. At the top, there is a search bar with the text 'This repository Search'. Below the search bar, the repository name 'Islandora / islandora_oai' is displayed, along with the text 'forked from adam-vessey/islandora_oai'. The main heading 'Islandora OAI' is prominently displayed, followed by a green badge that says 'build passing'.

Maintainers/Sponsors

Current maintainers:

- [Jordan Dukart](#)

Question Period

Question 1

Giancarlo and Nick, tell us about the impacts of harvesting your repositories.

How has it impacted contributors and users?

Question 2

Jordan, we heard the Islandora OAI module is has been re-worked and contributed to by 18 different people.

Tell us about the interest surrounding this module?

Question 3

Giancarlo and Nick, what sort of work did you have to do to model your data for harvesting?

Does it depend on the aggregator?

Question 4

Jordan, what is the configuration process like? What skills do you need to be able to complete it?

What are the most important elements?

Are there any tips or tricks you can share?

Question 5

Giancarlo and Nick, what steps would you recommend to those who want to harvest to aggregators?

Question 6

Jordan, what are some edge cases users should be aware of?

When would a project require customization to this module?

What to learn more?

Attend the
OAI-PMH in Islandora workshop
on Wednesday August 5, 2015
from 1 - 2:30 pm

For More Information

For more information about Islandora OAI see:

- Open Archives Initiative, <https://www.openarchives.org/pmh/>
- Islandora Online Discussion Forums, <https://groups.google.com/forum/#!searchin/islandora/OAI>
- Islandora Documentation, <https://wiki.duraspace.org/display/ISLANDORA715/Islandora+OAI>

For More Information

Examples of Free Service providers (harvester) include:

- WorldCat <http://www.worldcat.org>
- Opendoar <http://www.opendoar.org/>
- Europeana <http://www.europeana.eu/portal/>
- Culturaitalia <http://www.culturaitalia.it/> (Italy)
- DPLA <http://dp.la/info/hubs/become-a-hub/> (United States)

Thank you!

Moderator:

Erin Tripp, discoverygarden

erin@discoverygarden.ca

[@eeohalloran](#)

Panelist:

Giancarlo Birello, CNR-IRCrES

giancarlo.birello@ircres.cnr.it

Panelist:

Jordan Dukart, discoverygarden

jordan@discoverygarden.ca

[@discgarden](#)

Panelist:

Nick Ruest, York University

ruestn@yorku.ca

[@ruebot](#)