

OpenAIRE Advance Aggregator-as-a-Service with Canadian repositories

**OpenAIRE Task Group
CARL Open Repositories Working Group)**

Pierre Lasou, Université Laval

Kathleen Shearer, COAR

Michael Vandenburg, Queen's University

Mita Williams, University of Windsor

Geoff Harder, University of Alberta

Leah Vanderjagt, University of Alberta

Eugene Barsky, University of British Columbia

2019-06-13

1. Purpose

This plan defines the main characteristics and identifies the activities, schedule, and deliverables of a pilot project for the OpenAIRE Advance Aggregator-as-a-Service project with Canadian repositories.

The OpenAIRE Aggregator-as-a-Service aims to 1) Provide a view/portal for Canadian content and support tracking of Canadian research outputs 2) enrich and enhance metadata based on automated inference from other external sources (e.g. Tri-agency public grants database, Cross-ref, PubMed Central). It will contribute to increase discoverability of repository's contents and provide a comprehensive view of OA in Canada as well as support tracking and monitoring of compliance with the Tri-agency OA policy.

2. Reference Documents

OpenAIRE [Guidelines for Literature Repository Managers v4](#)

ORWG [OpenAIRE Lite Guidelines for for Canadian repositories](#)

OpenAIRE [Guidelines for Data Archives](#)

ORWG OpenAIRE Task Group [Terms of references](#)

OpenAIRE [content acquisition policies](#)

OpenAIRE [Terms of Agreement for Content Providers](#)

3. Pilot definition

The pilot objectives are :

- select canadian institutions willing to participate and commit to developments to ensure compliance.
- make necessary developments in repository software systems to allow them to be compliant with OpenAIRE Guidelines for Literature Repository Managers v4 Lite for Canada
- evaluate OpenAIRE capacity to identify funding information through data mining.

For literature guidelines, compliance will be tested :

- on a variety of software repositories (Eprints, DSpace, Bepress, Islandora, etc.).
- on PKP Open Journal Systems (OJS) platform using the new [OpenAIRE-plugin](#) that delivers the metadata in JATS.

For research data, compliance will be provided through the Portage discovery layer (FRDR, Dataverse North).

The pilot will be conducted during summer/fall 2019 (July to November).

4. Success criteria

All institutions participating in the pilot will be harvested by OpenAIRE and retrievable on OpenAIRE Search.

Funding information can be retrieved from Canadian public grants database using OpenAIRE text and data mining techniques.

5. Project Plan

Overview

To comply with OpenAIRE guidelines, one will need to use the new OpenAIRE specific metadata format in OAI PMH.

For compliance with the version 4 literature guidelines, the following OpenAIRE elements will be considered, which include only the mandatory elements (M = Mandatory ; MA = Mandatory if applicable):

1. Title (M)
2. Creator (M)
3. Contributor (MA)
4. FundingReference (MA)
7. Embargo Period Date (MA)
8. Language (MA)
9. Publisher (MA)
10. Publication Date (M)
11. Resource Type (M)
12. Description (MA)
- 14 Resource Identifier (M)
15. Access Rights (M)
17. Subject (MA)
23. File Location (MA)

Once a repository is harvested, it will be made available in the OpenAIRE knowledge graph and eventually through a specific Canadian view. OpenAIRE will text and data mine the Tri-agencies public grants databases to identify the funder name, which will become the “FundingReference” element in the record metadata.

All development during the pilot should be as generic as possible so that it can be generalized and shared with other Canadian institution.

Description of tasks

The pilot project will be undertaken in the following steps:

Step 1: Compliance with guidelines (July-August)

- Develop programs necessary to transform software metadata in OpenAIRE OAI format (XSL programs, etc.)
- Conduct tests
 - Test all access rights cases: OA, embargo, restricted
 - Test different publication type: article, thesis, chapter

Step 2: Validation (August-September)

- Validation of each participating institution needs to be conducted using OpenAIRE online validator.
- Test implementation and correct all errors if any

Step 3: Testing and debriefing (October-November)

- Test data mining for Tri agency funding, and inferences related to funder references
- Test search of content on OpenAIRE Canadian portal
- Produce a report on pilot experience

Participants

Participants in the pilot must agree to comply with OpenAIRE content acquisition policies and OpenAIRE Terms of Agreement for Content Providers.

Participating institutions: *TBD*

Resources

The pilot will require human resource time (especially librarian and developer time) from participating institutions.