

# OAI-PMH Metadata Harvesting

ARDCOAI-PMH = Open Archives Initiative Protocol for Metadata Harvesting

## Underlying Technology Overview

### Protocol to transfer metadata: HTTP

- request and response are sent via HTTP protocol
- requests are encoded as GET/POST operations
- responses are well-formed XML documents

### Metadata format

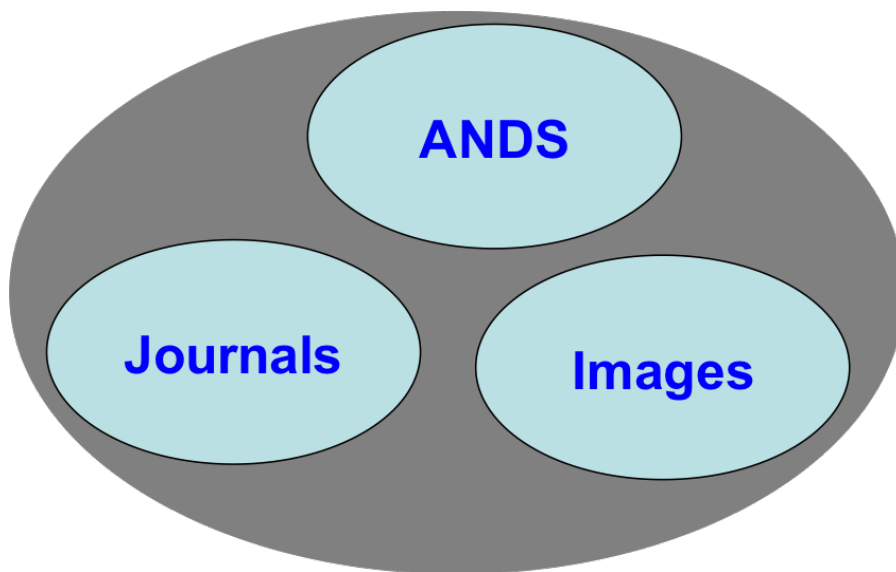
- any formats (e.g. Dublin Core, MARC, RIF-CS)
- default supported: Dublin Core
- ARDC supported: RIF-CS, ISO19115

## Data Provider vs Service Provider

OAI divides the world between data providers and service providers:

- Data Providers (Repositories)
  - expose metadata to harvesters
  - provide free access to metadata
- Service Providers (Harvesters)
  - client applications that issue OAI-PMH requests.

## Sets



- Allow for harvesting of sub-collections
- Sets can overlap: 1 item in multiple sets

## Overview of OAI-PMH verbs

### verb=Identify

Description of Institutional Repository

- Purpose: Return general information about institutional repository and its policies
- Parameters: None
- Sample URL: <https://researchdata.edu.au/registry/services/oai?verb=Identify>

```

<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2017-10-19T00:28:20Z</responseDate>
  <request verb="Identify">
https://researchdata.edu.au/registry/registry/services/oai
</request>
  <Identify>
    <repositoryName>Australian National Data Service (ANDS)</repositoryName>
    <baseUrl>https://researchdata.edu.au/registry/</baseUrl>
    <protocolVersion>2.0</protocolVersion>
    <earliestDatestamp>2010-01-12T05:03:58Z</earliestDatestamp>
    <deletedRecord>transient</deletedRecord>
    <granularity>YYYY-MM-DDThh:mm:ssZ</granularity>
    <adminEmail>services@ands.org.au</adminEmail>
  </Identify>
</OAI-PMH>

```

## verb=ListSets

Sets defined by Institutional Repository

- Purpose: Provide a hierarchical listing of sets in which records may be organised
- Parameters: None
- Sample URL: <https://researchdata.edu.au/registry/services/oai?verb=ListSets>

```

<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2017-10-19T00:40:58Z</responseDate>
  <request verb="ListSets">
https://researchdata.edu.au/registry/registry/services/oai
</request>
  <ListSets>
    <set>
      <setSpec>
datasource:Housing-and-building-National-Research-Center
</setSpec>
      <setName>Housing and building National Research Center</setName>
    </set>
    <set>
      <setSpec>datasource:Federal-University-of-Sergipe</setSpec>
      <setName>Federal University of Sergipe</setName>
    </set>
    <set>
      <setSpec>
datasource:DataSource-Example-with-4-Registry-Objects
</setSpec>
      <setName>DataSource Example with 4 Registry Objects</setName>
    </set>
    <set>
      <setSpec>datasource:Australian-National-University</setSpec>
      <setName>Australian National University</setName>
    </set>
    <set>
      <setSpec>datasource:IGO</setSpec>
      <setName>IGO</setName>
    </set>
    <set>
      <setSpec>datasource:testDStestDS</setSpec>
      <setName>testDStestDS</setName>
    </set>
    <set>
      <setSpec>datasource:ANDS</setSpec>
      <setName>ANDS</setName>
    </set>
  </ListSets>
</OAI-PMH>

```

## verb=ListMetadataFormats

Metadata formats supported by Institutional Repository

- Purpose: List metadata formats supported by institutional repository as well as their schema locations and namespaces
- Parameters: identifier – for a specific record (Optional)
- Sample URL: <https://researchdata.edu.au/registry/services/oai?verb=ListMetadataFormat>

```
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2017-10-19T00:43:25Z</responseDate>
  <request verb="ListMetadataFormats">
    https://researchdata.edu.au/registry/registry/services/oai
  </request>
  <ListMetadataFormats>
    <metadataFormat>
      <metadataPrefix>dc</metadataPrefix>
      <schema />
      <metadataNamespace />
    </metadataFormat>
    <metadataFormat>
      <metadataPrefix>scholix</metadataPrefix>
      <schema />
      <metadataNamespace />
    </metadataFormat>
    <metadataFormat>
      <metadataPrefix>oai_dc</metadataPrefix>
      <schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd</schema>
      <metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/</metadataNamespace>
    </metadataFormat>
    <metadataFormat>
      <metadataPrefix>rif</metadataPrefix>
      <schema>
        http://services.ands.org.au/documentation/rifcs/1.3/schema/registryObjects.xsd
      </schema>
      <metadataNamespace>
        http://ands.org.au/standards/rif-cs/registryObjects
      </metadataNamespace>
    </metadataFormat>
    <metadataFormat>
      <metadataPrefix>extRif</metadataPrefix>
      <schema />
      <metadataNamespace>
        http://ands.org.au/standards/rif-cs/extendedRegistryObjects
      </metadataNamespace>
    </metadataFormat>
  </ListMetadataFormats>
</OAI-PMH>
```

## verb=ListIdentifiers

OAI unique ids contained in Institutional Repository

- Purpose: List all unique identifiers corresponding to records in Institutional Repository
- Parameters:
  - from – start date (Optional)
  - until – end date (Optional)
  - set – set to harvest from (Optional)
- Sample URL: <https://researchdata.edu.au/registry/services/oai?verb=ListIdentifiers&metadataPrefix=rif>

```

<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2017-10-19T01:23:33Z</responseDate>
  <request verb="ListIdentifiers" metadataPrefix="rif">
https://researchdata.edu.au/registry/registry/services/oai
</request>
  <ListIdentifiers>
    <header>
      <identifier>oai:ands.org.au:1840</identifier>
      <datestamp>2014-02-28T02:30:53Z</datestamp>
      <setSpec>datasource:University-of-Sydney-AP05</setSpec>
      <setSpec>class:service</setSpec>
      <setSpec>group:The0x20University0x20of0x20Sydney</setSpec>
    </header>
    <header>
      <identifier>oai:ands.org.au:1842</identifier>
      <datestamp>2013-07-30T04:23:36Z</datestamp>
      <setSpec>datasource:Founders-and-Survivors-Genealogical-Connections</setSpec>
      <setSpec>class:service</setSpec>
      <setSpec>group:The0x20University0x20of0x20Melbourne</setSpec>
    </header>
    <header>
      <identifier>oai:ands.org.au:1849</identifier>
      <datestamp>2014-03-28T05:42:13Z</datestamp>
      <setSpec>datasource:AP34-CSIRO-SEQITOR</setSpec>
      <setSpec>class:service</setSpec>
      <setSpec>group:Commonwealth0x20Scientific0x20and0x20Industrial0x20Research0x20Organisation<
/setSpec>
    </header>
    <header>
      <identifier>oai:ands.org.au:2008</identifier>
      <datestamp>2012-08-16T00:43:39Z</datestamp>
      <setSpec>datasource:AusNC</setSpec>
      <setSpec>class:collection</setSpec>
      <setSpec>group:Australian0x20National0x20Corpus</setSpec>
    </header>
    .....
  </ListIdentifiers>
</OAI-PMH>

```

## verb=ListRecords

Listing of N records in Institutional Repository

- Purpose: Retrieves metadata for multiple records
- Parameters:
  - metadataPrefix – metadata format (Required)
  - from – start date (Optional)
  - until – end date (Optional)
  - set – set to harvest from (Optional)
- Sample URL: <https://researchdata.edu.au/registry/services/oai?verb=ListRecords&metadataPrefix=rif>

```

<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2017-10-19T01:28:55Z</responseDate>
  <request verb="ListRecords" metadataPrefix="rif">
https://researchdata.edu.au/registry/registry/services/oai
</request>
  <ListRecords>
    <record>
      <header>
        <identifier>oai:ands.org.au:1840</identifier>
        <datestamp>2014-02-28T02:30:53Z</datestamp>
        <setSpec>datasource:University-of-Sydney-AP05</setSpec>
        <setSpec>class:service</setSpec>
        <setSpec>group:The0x20University0x20of0x20Sydney</setSpec>
      </header>

```

```

<metadata>
  <registryObjects xmlns="http://ands.org.au/standards/rif-cs/registryObjects" xmlns:xsi="
http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://ands.org.au/standards/rif-cs
/registryObjects http://services.ands.org.au/documentation/rifcs/schema/registryObjects.xsd">
    <registryObject group="The University of Sydney">
      <key>AP05_USyd_thisshouldbeautogenerated</key>
      <originatingSource type="">
http://researchdata.ands.org.au/registry/orca/register_my_data
</originatingSource>
        <service type="search-http" dateModified="2012-10-23T00:00:00Z">
          <name type="primary">
            <namePart>
CATAMI: Collaborative and Annotation Tools for Analysis of Marine Imagery and Video
</namePart>
              </name>
              <description type="full">
<p>&lt;/p>The CATAMI web site provides a location of the deposit and access of various underwater imagery,
including data from Baited Remote Underwater Video (BRUV), Autonomous Underwater Vehicles (AUV), Diver
Operated Video (DOV) and Towed Imagery (TI). Functionality and uses of this service include:&lt;/p>
<ul style="list-style-type: circle;">
<li>&lt;/li>Methods for querying the image database based on criteria
such as depth, position, time, image labels, etc. will allow relevant subsets of the data to be quickly
extracted from the repository for further analysis;&lt;/li>
<li>&lt;/li>Online access to geotiff images of
the integrated terrain models that allow broad scale patterns to be examined and allows users to identify
particular subsets of the data that are of interest.&lt;/li>
<li>&lt;/li>Access to video data collected by
BRUVS, UTV and ROV systems. &lt;/li>
<li>&lt;/li>Presentation of summary statistics for dives and campaigns
based on analysis output. The ability to quickly examine the makeup of individual dives, including sample
images of the relevant habitat types, will allow end users to quickly examine and understand the content of
individual deployments.&lt;/li>
</ul>
</description>
                <description type="deliverymethod">&lt;/p>software&lt;/p></description>
                <description type="note">
<div>&lt;/div>This project includes development funded by the Australian National Data Service (ANDS, &lt;a
href="http://ands.org.au">http://ands.org.au&lt;/a>) and the National eResearch Collaboration Tools
and Resources (NeCTAR, &lt;a href="http://nectar.org.au">http://nectar.org.au&lt;/a>).&lt;/div>
</div>
<div>&lt;/div>ANDS is supported by the Australian Government through the National
Collaborative Research Infrastructure Strategy Program and the Education Investment Fund (EIF) Super Science
Initiative.&lt;/div>
<div>&lt;/div> NeCTAR is an Australian Government project
conducted as part of the Super Science initiative and financed by the Education Investment Fund. The
University of Melbourne has been appointed the lead agent by the Commonwealth of Australia, Department of
Industry, Innovation, Science, Research and Tertiary Education.&lt;/div>
</description>
          <rights>
            <rightsStatement rightsUri="">Rights accessible via the associated license.<
/rightsStatement>
            <licence type="CC-BY" rightsUri="http://creativecommons.org/licenses/by/3.0
/" />
            <accessRights type="" rightsUri="">Open access.</accessRights>
          </rights>
          <identifier type="uri">http://catami.org</identifier>
          <location>
            <address>
              <electronic type="url">
                <value>http://catami.org</value>
              </electronic>
            </address>
          </location>
          <relatedObject>
            <key>AODN:ae70eb18-b1f0-4012-8d62-b03daf99f7f2</key>
            <relation type="makesAvailable" />
          </relatedObject>
          <relatedObject>
            <key>AODN:f47a6929-1f19-4724-b74b-7c8579872cb7</key>
            <relation type="makesAvailable" />
          </relatedObject>
          <relatedObject>
            <key>AODN:783f9b8d-4ce2-417b-aeb1-234799dc4696</key>
            <relation type="makesAvailable" />
          </relatedObject>
          <relatedObject>
            <key>AODN:stefanw@acfr.usyd.edu.au</key>
            <relation type="hasAssociationWith" />

```

```

        </relatedObject>
        <relatedObject>
            <key>AODN:6be08828-275d-4a5c-9478-f5fe20613709</key>
            <relation type="makesAvailable" />
        </relatedObject>
        <relatedObject>
            <key>AODN:0f379372-f0c3-4c86-aa25-4aeaa3986635</key>
            <relation type="makesAvailable" />
        </relatedObject>
        <subject type="anzsrc">0600</subject>
        <subject type="local">Oceans</subject>
        <subject type="local">Marine</subject>
        <relatedInfo type="website">
            <identifier type="uri">http://catami-australia.blogspot.com.au/</identifier>
        </relatedInfo>
    </service>
</registryObject>
</registryObjects>
</metadata>
</record>
</ListRecords>
.....
</OAI-PMH>

```

## verb=GetRecord

Listing of a single record from Institutional Repository

- Purpose: Returns the metadata for a single identifier in the form of an OAI record
- Parameters:
  - identifier – unique id for record (Required)
  - metadataPrefix – metadata format (Required)
- Sample URL: <https://researchdata.edu.au/registry/services/oai?verb=GetRecord&metadataPrefix=rif&identifier=http://nla.gov.au/nla.party-1508909>

```

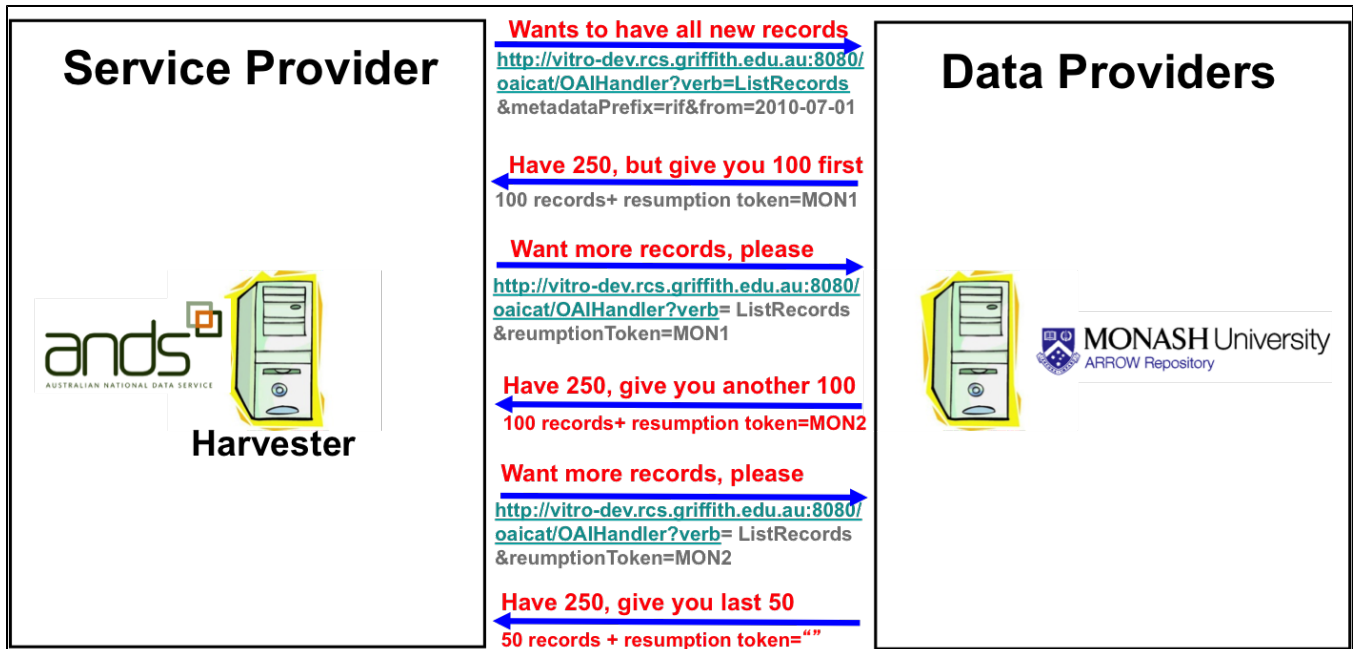
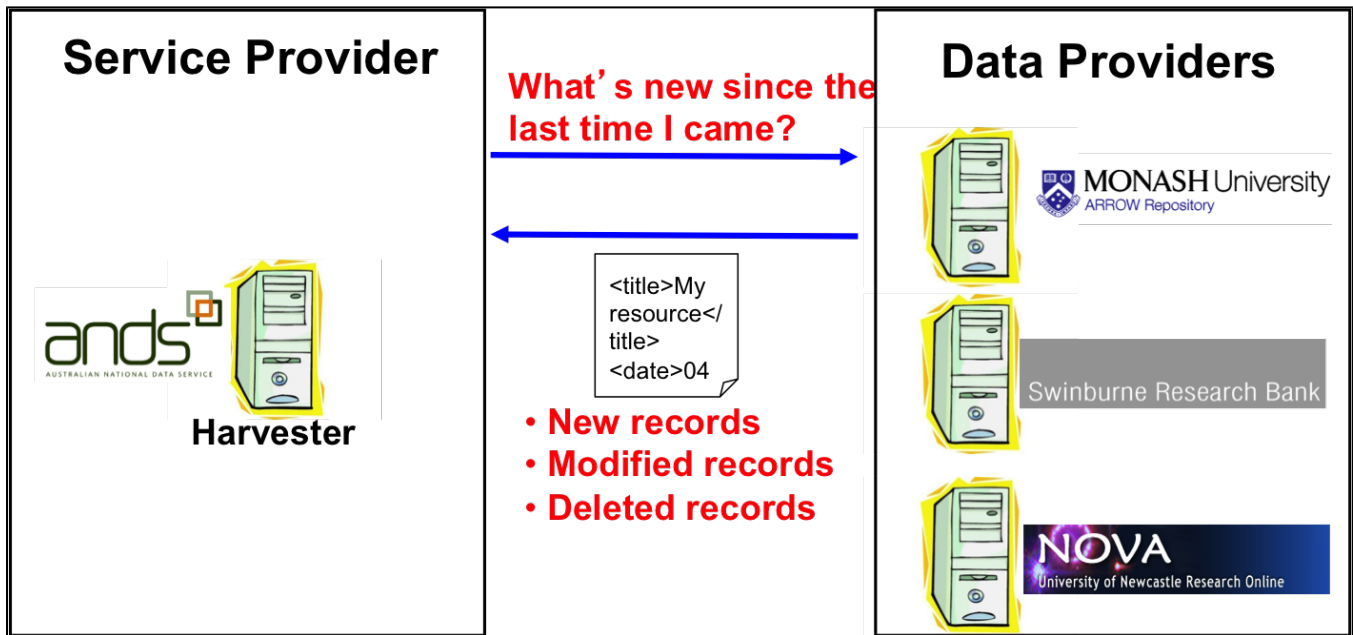
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2017-10-19T01:35:28Z</responseDate>
  <request verb="GetRecord" metadataPrefix="rif" identifier="http://nla.gov.au/nla.party-1508909">
https://researchdata.edu.au/registry/registry/services/oai
</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:ands.org.au:671935</identifier>
        <datestamp>2016-06-29T17:13:39Z</datestamp>
      </header>
      <metadata>
        <registryObjects xmlns="http://ands.org.au/standards/rif-cs/registryObjects" xmlns:extRif="
http://ands.org.au/standards/rif-cs/extendedRegistryObjects" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://ands.org.au/standards/rif-cs/registryObjects http://services.ands.org.au
/documentation/rifcs/schema/registryObjects.xsd">
          <registryObject group="National Library of Australia">
            <key>http://nla.gov.au/nla.party-1508909</key>
            <originatingSource type="authoritative">http://trove.nla.gov.au/people<
/originatingSource>
              <party dateModified="2015-01-12T04:59:23Z" type="group">
                <identifier type="AU-ANL:PEAU">http://nla.gov.au/nla.party-1508909</identifier>
                <identifier type="AU-VU:EOAS">http://www.eoas.info/biogs/P005286b.htm<
/identifier>
                  <name type="primary">
                    <namePart>Australian National Data Service</namePart>
                  </name>
                  <location>
                    <address>
                      <electronic type="url">
                        <value>http://nla.gov.au/nla.party-1508909</value>
                      </electronic>
                    </address>
                  </location>
                  <description type="brief">The Australian National Data Service (ANDS) was
created in 2009 with the aim to provide a cohesive collection of research resources from all research
institutions to improve access and use of Australia's research data collections.</description>
                </party>
              </registryObject>
            </registryObjects>
          </metadata>
        </record>
      </GetRecord>
    </OAI-PMH>

```

## Datestamp

- Each record needs a datestamp:
  - date of creation
  - date of modification.
- Dates are used to allow:
  - harvesting by date range
  - support for incremental harvesting.

## Incremental harvesting



### Existing OAI-PMH solutions

- jOAI
- OAI Cat
- Proai for Fedora

### ARDC metadata and transforms

#### ARDC harvester supported metadata format

- RIF-CS (ISO 2146)
- ISO 19115 (Geographic Information-Metadata)
- ISO 19139 (Marine Community Profile)

#### Transforms (XSLT)



- [DC to RIF-CS](#)
- [MARCS to RIF-CS](#)
- [MODS to RIF-CS](#)

More information is available from the [Crosswalks: Transform your metadata](#) page.

[« Research Data Australia \(RDA\) Registry](#) [Configuring jOAI to support RIF-CS](#) »