Technical Development

UK Research Data Registry and Discovery Service Workshop
London, 16 June 2014

Choosing a platform

- Pilot with limited time and resources
  - **Use existing software**
- Is anyone running a (national) research data registry elsewhere?
  - **ANDS: http://researchdata.ands.org.au**
- Is their software readily available under an open-source licence?
  - **Yes: https://github.com/au-research/ (Apache License 2.0)**

Research Data Australia

Research Data Australia

About Collections Parties Activities Services Themes

Search for Research Data

Browse by Subject Area

Browse by Map Coverage

Advanced Search

What's in Research Data Australia

Collections (97690)
Research datasets or collections of research materials.

Parties (25659)
Researchers or research organisations that create or maintain research datasets or collections.

Activities (40695)
Projects or programs that create research datasets or collections.

Services (182)

Spotlight on research data

Professor Matthew England

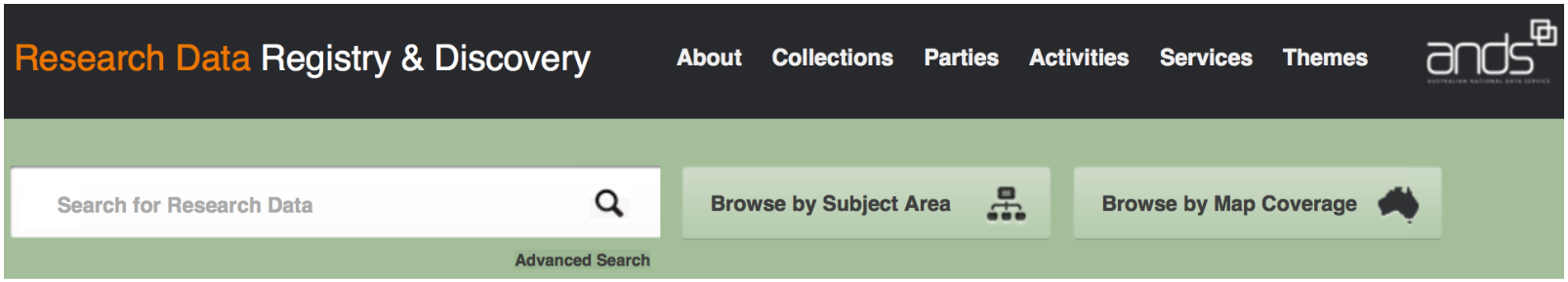
The Australian Academy of Science has recently elected as a Fellow, Professor England of the Climate Change Research Centre at the University of New South Wales, for his work on modelling the Southern Ocean and deep ocean ventilation, and its application to climate models. Professor England's research interests relate to the global-scale ocean circulation and its influence on regional climate. Using ocean and coupled climate models in combination with observations, he can study what controls ocean currents and how these currents affect climate and climate variability on time-scales of seasons to centuries.

Discover collections contributed by Professsor England in Research Data Australia >>>

Installation

- CentOS Linux instance in Microsoft's Azure cloud platform
  - Thanks, Microsoft Research!
- PHP application
- MySQL database
- Apache Solr indexing
- Harvester is separate Java component
- Few issues getting things up and running
  - Some help from the RDA with the Harvester
  - Some performance issues

UK Research Data Registry and Discovery Service: http://rdrds.cloudapp.net



What's in the Research Data Registry and Discovery Service

Collections (49)
Research datasets or collections of research materials.

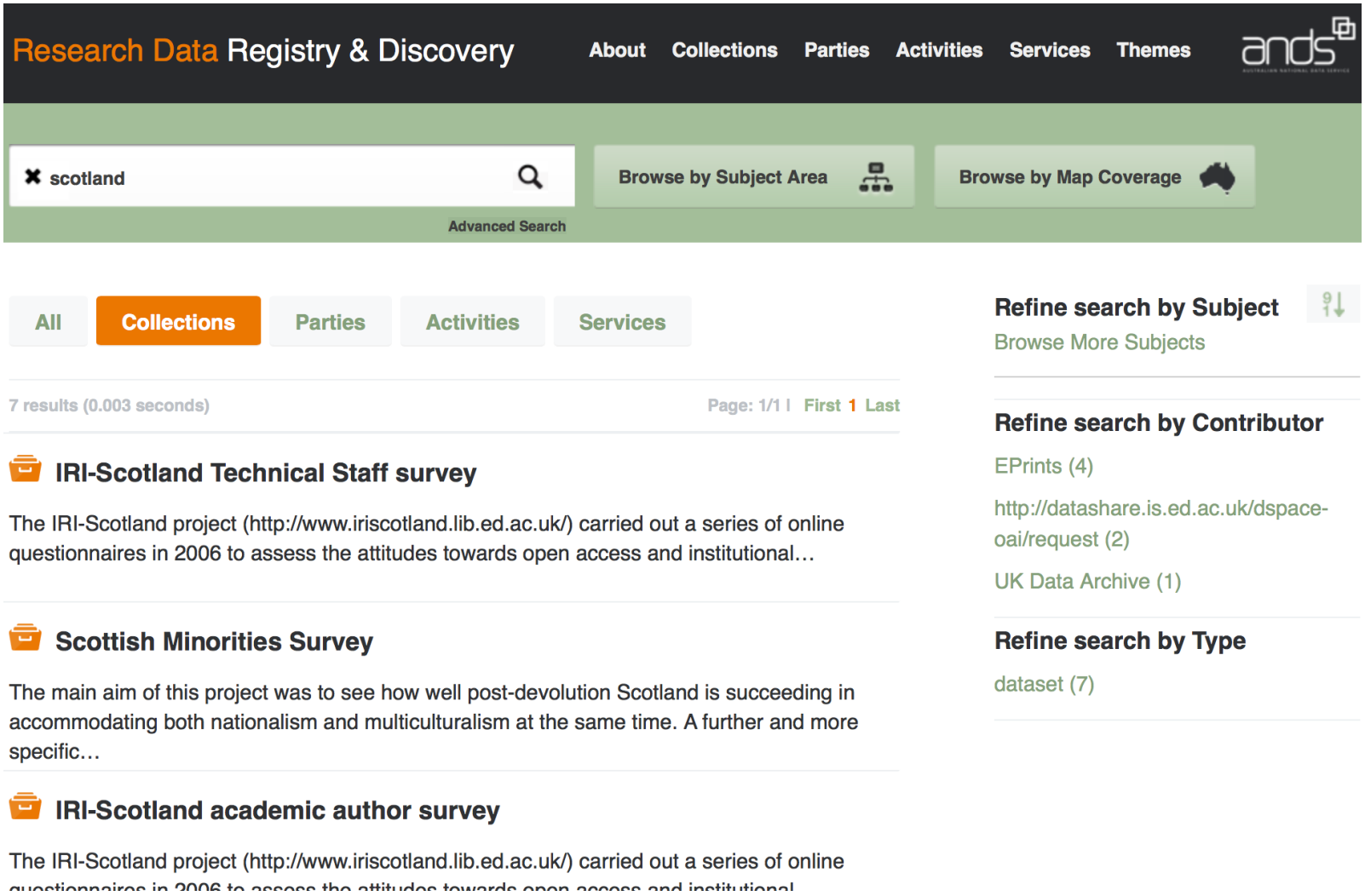
Parties (36)
Researchers or research organisations that create or maintain research datasets or collections.

Activities (0)
Projects or programs that create research datasets or collections.

Spotlight on research data

Who contributes to the Research Data Registry and Discovery Service?

UK Research Data Registry and Discovery Service: http://rdrds.cloudapp.net



UK Research Data Registry and Discovery Service: http://rdrds.cloudapp.net

IRI-Scotland Technical Staff survey

The IRI-Scotland project (http://www.iriscotland.lib.ed.ac.uk/) carried out a series of online questionnaires in 2006 to assess the attitudes towards open access and institutional repositories within the higher education community in Scotland. In total, three questionnaires were targeted at different stakeholder groups within the community - academic authors, technical staff responsible for repository development, and senior management from academic libraries.

The second IRI-Scotland survey was targeted at technical staff, usually based in academic libraries or aligned information services support groups, who would be responsible for developing a digital repository in their institution. The questions were aimed to help determine the functional requirements needed to build a national hosted repository service that would be suitable for the current and future repository infrastructure in Scotland.

We present here anonymous data from the technical staff survey in comma separated value format.

How to Cite this Collection

Citation (Metadata):

Greig, Morag Ashworth, Susan Andrew, Theo (2008,2008): IRI-Scotland Technical Staff survey.

Identifiers

Local: Andrew, Theo; Greig, Morag; Ashworth, Susan. (2008). IRI-Scotland Technical Staff survey [Dataset].

Local: http://hdl.handle.net/10283/9

Access

Connections

People

Susan Ashworth (PI)

Theo Andrew (PI)

Morag Greig (PI)

Suggested Links

Internal Records

1 record with matching subjects

External Records

48 records from DataCite

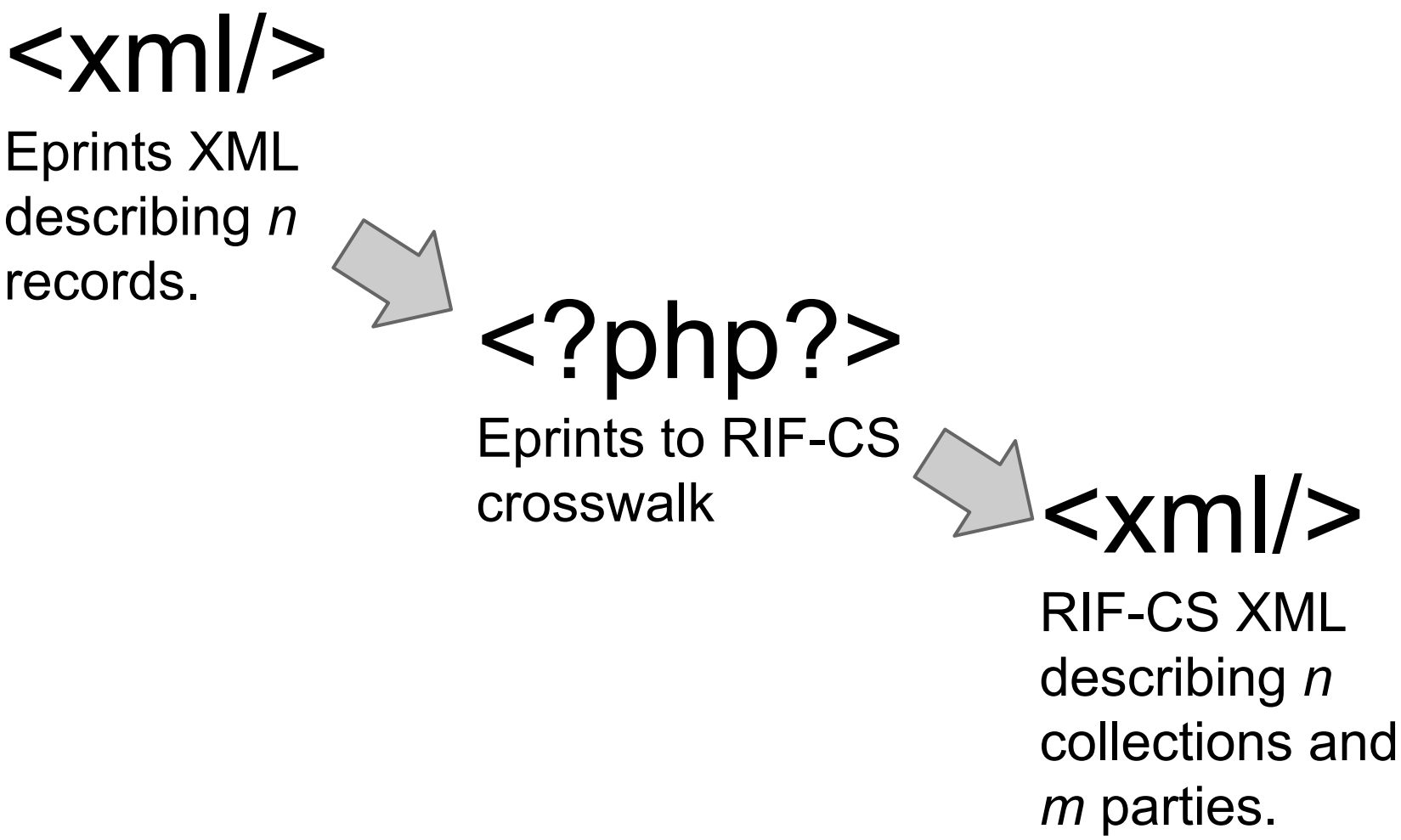
Crosswalks

- Imported metadata has to be converted to the RIF-CS format used by Research Data Australia
- RIF-CS: Registry Interchange Format - Collections and Services
  - Based on ISO 2146 - an information model for "registry services for libraries and related organisations"

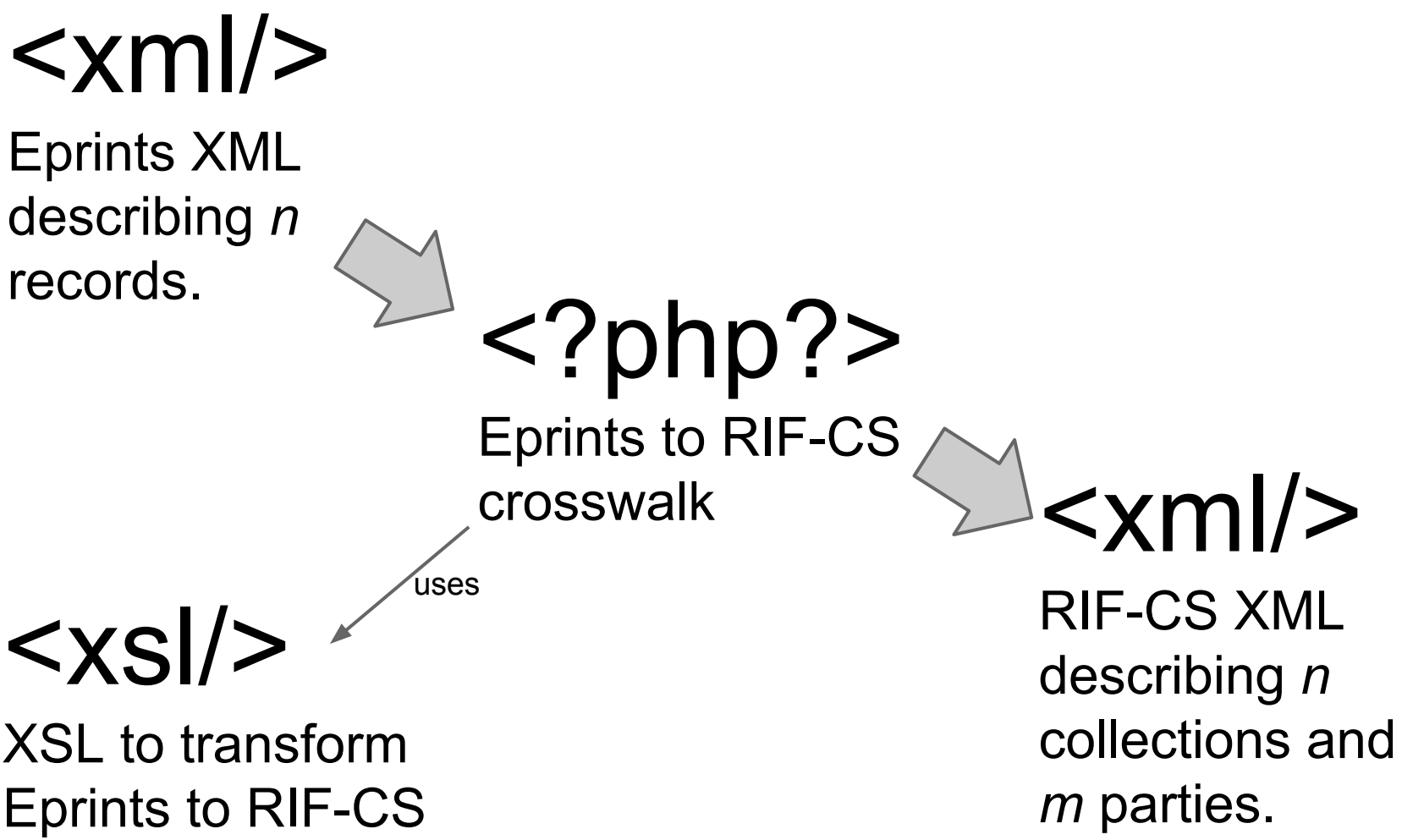
Crosswalks

- RIF-CS (and ISO 2146) describes these entities or *registry objects* and the relationships between them:
  - Collections (e.g. of data)
  - Parties (people or organisations)
  - Services
  - Activities (grants, projects, etc.)

Crosswalks

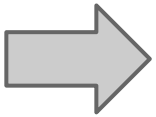

<xml/>
Eprints XML describing *n* records.
<?php?>
Eprints to RIF-CS crosswalk
<xml/>
RIF-CS XML describing *n* collections and *m* parties.

Crosswalks


<xml/>
Eprints XML describing *n* records.
<?php?>
Eprints to RIF-CS crosswalk
uses
<xsl/>
XSL to transform Eprints to RIF-CS
<xml/>
RIF-CS XML describing *n* collections and *m* parties.

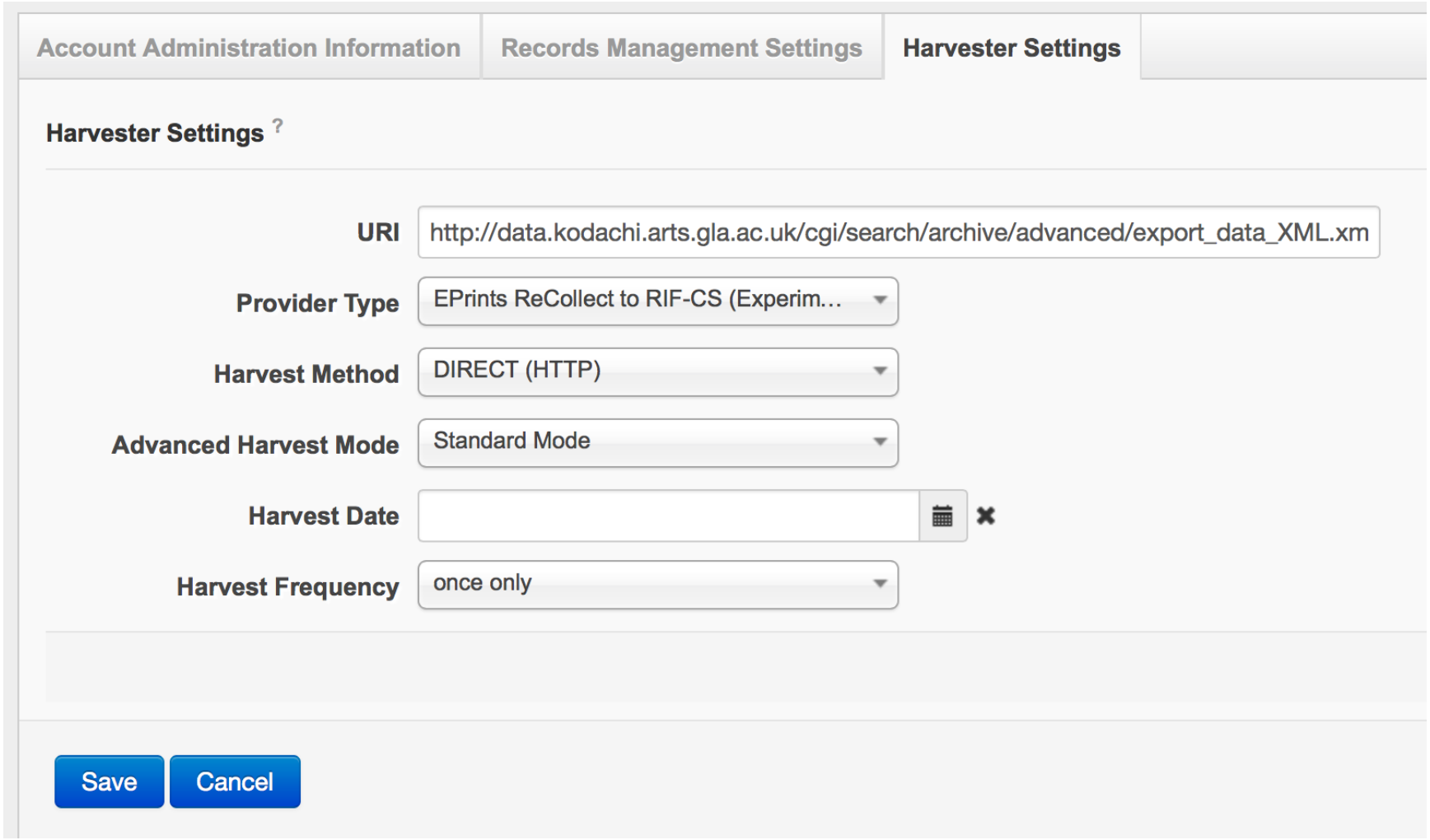
Crosswalks

```
<records>
  <record>
    <identifier>[DOI 1]</identifier>
    <author>Joe Bloggs</author>
  </record>
  <record>
    <identifier>[DOI 2]</identifier>
    <author>Joe Bloggs</author>
  <record>
</records>
```

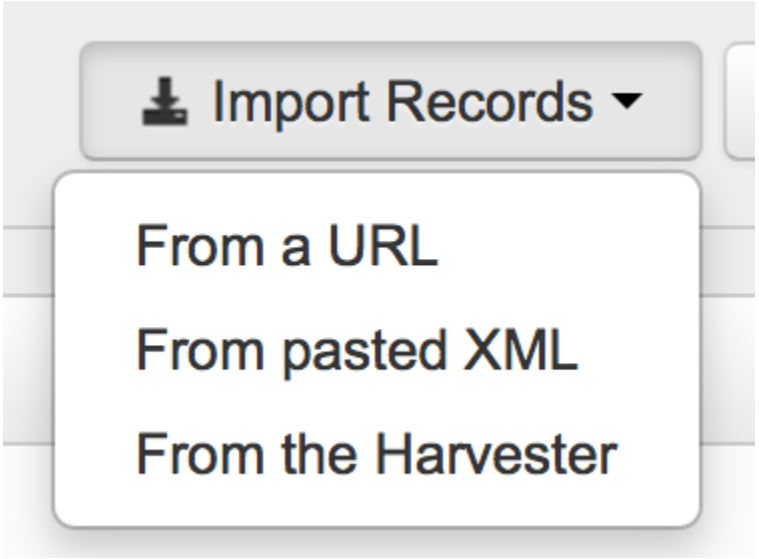


```
<registryObjects>
  <registryObject type='collection'>
    <identifier>[DOI 1]</identifier>
    <relatedTo type='author'>
      [auto-gen ID 1]
    </relatedTo>
  <registryObject>
  <registryObject type='party'>
    <identifier>[auto-gen ID 1]</identifier>
    <name>Joe Bloggs</name>
  <registryObject>
  <registryObject type='collection'>
    <identifier>[DOI 2]</identifier>
    <relatedTo type='author'>
      [auto-gen ID 2]
    </relatedTo>
  <registryObject>
  <registryObject type='party'>
    <identifier>[auto-gen ID 2]</identifier>
    <name>Joe Bloggs</name>
  <registryObject>
</registryObjects>
```

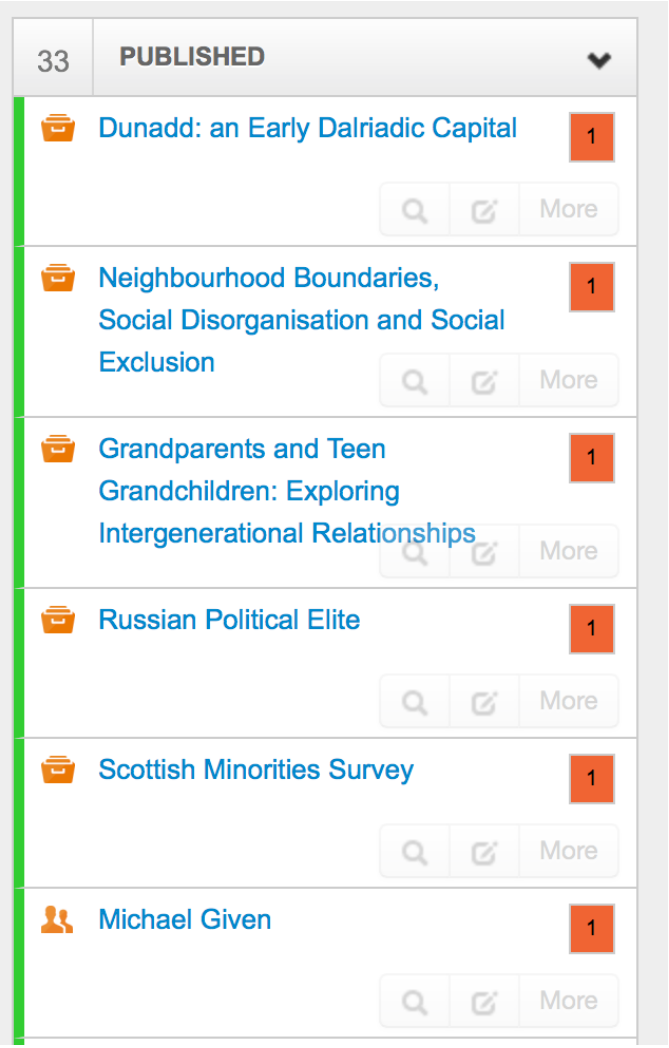
Harvesting


Account Administration Information
Records Management Settings
Harvester Settings
Harvester Settings ?
URI
http://data.kodachi.arts.gla.ac.uk/cgi/search/archive/advanced/export_data_XML.xm
Provider Type
EPrints ReCollect to RIF-CS (Experim…
Harvest Method
DIRECT (HTTP)
Advanced Harvest Mode
Standard Mode
Harvest Date
Harvest Frequency
once only
Save
Cancel

Importing


Import Records
From a URL
From pasted XML
From the Harvester

Imported Records



- Imported records are “submitted for assessment”
- Moved through “assessment in progress” to “published”
- Records given a quality level of 1-3

Metadata Quality Levels

1. [Includes all] Required RIF-CS Schema Elements
2. [Meets] Required Metadata Content Requirements
3. [Meets] Recommended Metadata Content Requirements

- Records can be imported successfully while failing to meet level 1 requirements

http://ands.org.au/resource/metadata-content-requirements.html

Metadata Quality

- Unique Identifiers - not just for datasets, but for related entities including individuals and organisations
  - ORCID?
  - Fundref?
- Look beyond Dublin Core
- Have seen records lacking basic information e.g. title

Improving Metadata

- Opportunity to coordinate and collaborate on
  - Selecting schemas
  - Improving quality
    - Detail
    - Unique Identifiers
- The fewer crosswalks to be maintained the better
  - Are subject-specific formats suitable for a cross-disciplinary registry?

Suitability of ANDS software

- Designed to be a Research Data Registry
- Open-source on GitHub
- Active development
  - Big changes coming to Harvester
  - Moves to reduce coupling to RIF-CS
- Designed by ANDS for use in their context
  - Not readily configurable for others
  - Modifying for use in UK context complicates implementation of future updates
- Currently coupled to RIF-CS

Possible way forward with RDA

- Configure/modify RDA for use in UK context
  - Will need substantial work to adapt future versions of RDA as they become available
- Collaborate with ANDS to turn RDA into an easily configurable product, readily deployable in a range of contexts
  - Significant effort, but of greater value to the wider community

Alternatives

- ~~Develop new registry application from scratch~~
- Use other existing platform e.g. CKAN
  - Readily configured and deployed
  - Not designed to be research data registry specifically
  - Will require development of extensions likely to be of use to the wider community

Recommendations

- Evaluation of alternatives e.g. CKAN
- Effort to agree more broadly on metadata schemas
- If going with the ANDS software
  - Collaborate to
    - Make it more easily adapted to other contexts
    - Improve documentation
    - Shape future developments
    - Develop associated components (Harvester)
  - Further develop and test crosswalks

?