Running an Open Source National Research Data Registry

Amir Aryani
Australian National Data Service
What’s in Research Data Australia

Collections (102928)
Research datasets or collections of research materials.

Parties (26905)
Researchers or research organisations that create or maintain research datasets or collections.

Activities (42046)
Projects or programs that create research datasets or collections.

Spotlight on research data

CATAMI:
Collaborative and Annotation Tools for Analysis of Marine Imagery and Video

In recent years, ecologists and biologists have increasingly relied on remote video and image-based methods to examine and monitor marine habitats, particularly those beyond easy diving depth. These data streams are currently being collected by the end user community at sites around Australia using modern image and video based tools including Baited Remote Underwater Video Systems (BRUVS), Autonomous Underwater Vehicle (AUV), Remotely Operated Vehicle (ROV) and Underwater Towed Video (UTV) systems. Transforming raw visual data into quantitative information useful for science and policy decisions requires substantial effort by human experts. The University of Sydney has led a collaborative project to develop common labelling standards, and to transform and collate existing datasets that support marine research across Australia. The
Agenda

- Open source software
- Metadata
- Business of running a registry
What makes a good open source software package?
Make your code understandable / readable
Make your API easy to use

Controlled Vocabulary Widget

Instantly add Data Classification capabilities to your data capture tools

Learn More

researchdata.ands.org.au/developers
Community blog

Release 12.2 – Prefix Agnostic Harvester

Posted on April 22nd, 2014

Prior to Release 12.2 the only metadataPrefix for OAI was 'rif'. Even if coded with another prefix the harvester would over ride the provided prefix with 'rif'.

As of release 12.2 the metadataPrefix is no longer limited to 'rif'. The prefix that is required when a harvest is created can be configured in the global_config.php. This also allows the user to map prefixes to existing cross-walks. See example configuration code below.

```
// Provider types
// allow to add different metadataPrefix for OAI harvesting
// also can be configured to run cross walk if needed.

$ENV['provider_types'] = array(
    'RIF-CS' => array('prefix' => 'rif', 'cross_walk' => ''),
    'RIF-CS-2' => array('prefix' => 'rif2', 'cross_walk' => ''),
    'OAI-DC' => array('prefix' => 'oai_dc', 'cross_walk' => 'Dublin Core to RIFCS')
);
```

Once the configuration for a provider has been set the available metadata prefix type will be available for selection in the Data Source admin settings eg 'OAI-DC'. If your cross_walk is blank then no cross will be executed and will assume that the provided xml is RIF-CS. This will allow users to configure multiple metadata prefixes for RIF-CS.

Despite the fact that prefixes aren’t required to be passed to the provider when requesting a feed using the GET method it can still be used to identify cross-walks for GET feeds. This means that crosswalks can be defined for both GET and OAI-PMH feeds.

```
$ENV['provider_types'] = array(
    'CSV2RIF' => array('prefix' => 'csv', 'cross_walk' => 'CSV 2 RIF-CS')
);
```

This configuration compliments the existing crosswalk handler and crosswalks can still be added as a class into the crosswalks directory.

developers.ands.org.au/
Members of the community help each other
Software bugs are inevitable.
Software bugs are inevitable

How quickly can you deal with them?
Enabling creativity
Modular code makes it easy to
● change,
● debug,
● fix,

... and enables creativity.
What do I need to do to install this?  
... should I press <next>?
Do we have the right metadata schema?
RIF-CS: Registry Interchange Format - Collections and Services

What’s in Research Data Australia

- **Collections** (102928)
  Research datasets or collections of research materials.

- **Parties** (26905)
  Researchers or research organisations that create or maintain research datasets or collections.

- **Activities** (42046)
  Projects or programs that create research datasets or collections.

- **Services** (355)
  Services that support the creation or use of research datasets or collections.

Does your model work for everyone?
Versioning

RIF-CS 1.5.0:
- Schema Guidelines
- Overview Diagram
- RIF-CS Schema:
  - Registry Objects Schema
  - Activity Schema
  - Collection Schema
  - Party Schema
  - Service Schema
  - Type Definitions
- Controlled Vocabularies
- Schema Documentation

RIF-CS 1.4.0:
- Schema Guidelines
- Overview Diagram
- RIF-CS Schema:
  - Registry Objects Schema
  - Activity Schema
  - Collection Schema
  - Party Schema
  - Service Schema
  - Type Definitions
- Controlled Vocabularies
- Schema Documentation

RIF-CS 1.3.0:
- Schema Guidelines
- Overview Diagram
- RIF-CS Schema:
  - Registry Objects Schema
What are the top 3 things to know a national registry?
Working groups: collective decisions are more expensive but more sustainable
Tip: use an issue tracking software

![BSWG Issues By Status](image)

**Filter Results: BSWG Issues (Raised)**

<table>
<thead>
<tr>
<th>Assignee</th>
<th>Created</th>
<th>Key</th>
<th>Reporter</th>
<th>Status</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cel Pilapil</td>
<td>09/Sep/14</td>
<td>BSG-138</td>
<td>Cel Pilapil</td>
<td>RAISED</td>
<td>Project Falling</td>
</tr>
<tr>
<td>Cel Pilapil</td>
<td>29/Aug/14</td>
<td>BSG-137</td>
<td>Cel Pilapil</td>
<td>RAISED</td>
<td>Automatic commit on a contributed rif-cs.</td>
</tr>
<tr>
<td>Cel Pilapil</td>
<td>06/May/14</td>
<td>BSG-133</td>
<td>Cel Pilapil</td>
<td>RAISED</td>
<td>Enhancement selection facet</td>
</tr>
</tbody>
</table>

1–3 of 3

**Filter Results: BSWG Issues ("Analysis in Progress")**

<table>
<thead>
<tr>
<th>Assignee</th>
<th>Created</th>
<th>Key</th>
<th>Reporter</th>
<th>Status</th>
</tr>
</thead>
</table>

![ 따옴표 보른 액션 수준의 문제반의 표시 방법

While not shown in the image, the screenshot displays a dashboard titled "BSWG Issues By Status" with filters for "BSWG Issues (Raised)", "BSWG Issues ("Analysis in Progress")", and "BSWG Issues (Closed)". Each filter shows a table with columns for Assignee, Created, Key, Reporter, Status, and Summary. The table for "BSWG Issues (Raised)" includes three rows with the following details:

1. **Cel Pilapil**
   - Created: 09/Sep/14
   - Key: BSG-138
   - Reporter: Cel Pilapil
   - Status: RAISED
   - Summary: Project Falling

2. **Cel Pilapil**
   - Created: 29/Aug/14
   - Key: BSG-137
   - Reporter: Cel Pilapil
   - Status: RAISED
   - Summary: Automatic commit on a contributed rif-cs.

3. **Cel Pilapil**
   - Created: 06/May/14
   - Key: BSG-133
   - Reporter: Cel Pilapil
   - Status: RAISED
   - Summary: Enhancement selection facet

The table for "BSWG Issues ("Analysis in Progress")" includes a row with the following details:

- **Cel Pilapil**
  - Created: 06/May/14
  - Key: BSG-133
  - Reporter: Cel Pilapil
  - Status

The "BSWG Issues By Status" filter also includes tabs for "Director Dashboard", "Change Control", "My Work Board", and "Technical Lead's Dashboard".
It is important to be agile as every contributor can change their systems at any time.
Final remark: “Running a national research data registry is a collaborative effort.”
Attribution for the photos

Code creativity: [https://www.flickr.com/photos/txmx-2/6145817075/in/photolist-b7r6g-6fmKyW-bpd8Ht-dLUWMb-chYPBy-51p7BJ-3AvBlV-bchsp6-aodrmH-4nqiZn-an5Vin-98CdBL-e1JYE-9dUbHm-auWNV-5RUuNN-a28TP6-6NKzvV-78bnLW-4UIJRv-y9A4Eb-7d7ipd-9p77om-buRfLu-5WvBYq-aogX72-4SexF-8PNWFyeY6F-8dpeZZ-8dsrWJ-4mtkin-6Rh1p-vW8f0-3ijuU-fd3oYe-7q1fpaMGSS4-kmhEbc-jmr8PB-azQhijy-icMHWD-4mtkij6-6SgG7q-3hkoR7HjQa-dnQW4g-i3e7sw-fpreHD-5XzfW/](https://www.flickr.com/photos/txmx-2/6145817075/in/photolist-b7r6g-6fmKyW-bpd8Ht-dLUWMb-chYPBy-51p7BJ-3AvBlV-bchsp6-aodrmH-4nqiZn-an5Vin-98CdBL-e1JYE-9dUbHm-auWNV-5RUuNN-a28TP6-6NKzvV-78bnLW-4UIJRv-y9A4Eb-7d7ipd-9p77om-buRfLu-5WvBYq-aogX72-4SexF-8PNWFyeY6F-8dpeZZ-8dsrWJ-4mtkin-6Rh1p-vW8f0-3ijuU-fd3oYe-7q1fpaMGSS4-kmhEbc-jmr8PB-azQhijy-icMHWD-4mtkij6-6SgG7q-3hkoR7HjQa-dnQW4g-i3e7sw-fpreHD-5XzfW/)

Software bug: [https://www.flickr.com/photos/guitavares/1703252007/in/photolist-b7r6g-6fmKyW-bpd8Ht-dLUWMb-chYPBy-51p7BJ-3AvBlV-bchsp6-aodrmH-4nqiZn-an5Vin-98CdBL-e1JYE-9dUbHm-auWNV-5RUuNN-a28TP6-6NKzvV-78bnLW-4UIJRv-y9A4Eb-7d7ipd-9p77om-buRfLu-5WvBYq-aogX72-4SexF-8PNWFyeY6F-8dpeZZ-8dsrWJ-4mtkin-6Rh1p-vW8f0-3ijuU-fd3oYe-7q1fpaMGSS4-kmhEbc-jmr8PB-azQhijy-icMHWD-4mtkij6-6SgG7q-3hkoR7HjQa-dnQW4g-i3e7sw-fpreHD-5XzfW/](https://www.flickr.com/photos/guitavares/1703252007/in/photolist-b7r6g-6fmKyW-bpd8Ht-dLUWMb-chYPBy-51p7BJ-3AvBlV-bchsp6-aodrmH-4nqiZn-an5Vin-98CdBL-e1JYE-9dUbHm-auWNV-5RUuNN-a28TP6-6NKzvV-78bnLW-4UIJRv-y9A4Eb-7d7ipd-9p77om-buRfLu-5WvBYq-aogX72-4SexF-8PNWFyeY6F-8dpeZZ-8dsrWJ-4mtkin-6Rh1p-vW8f0-3ijuU-fd3oYe-7q1fpaMGSS4-kmhEbc-jmr8PB-azQhijy-icMHWD-4mtkij6-6SgG7q-3hkoR7HjQa-dnQW4g-i3e7sw-fpreHD-5XzfW/)

Community: [https://www.flickr.com/photos/purrr/126597849/in/photolist-cbR8x-iRL2Xc-cxhBwN-6Lof3x-njiND1-9SAJAG-K83wP-7ugbbU-BvU6G-6KjXVN-28VCp4-4fTR98-nhDCqi-4j8w4p-2TrnmRm-ikTP1f-SzsfQ-bkkuEB-8gcf1w-a26aqqX-5Q6Bq-8BPCEC-7R58b8-oqFte-bzy1PV-7Y7Ywwk-fGazWu-y59xh-hYY15-bzd5ym-5XudVL-cEwUnU-laf3S1-bFQKS6-njRt8-34Wub7-nAyTys-GyMse-3b8udo-hglxR5-5R5ebY-aZk2zz-2WxQ3n-7w9QMBeccf75F-4Fh97Y-3mzwqg-7yhraF-8PFLQb-qA3pC](https://www.flickr.com/photos/purrr/126597849/in/photolist-cbR8x-iRL2Xc-cxhBwN-6Lof3x-njiND1-9SAJAG-K83wP-7ugbbU-BvU6G-6KjXVN-28VCp4-4fTR98-nhDCqi-4j8w4p-2TrnmRm-ikTP1f-SzsfQ-bkkuEB-8gcf1w-a26aqqX-5Q6Bq-8BPCEC-7R58b8-oqFte-bzy1PV-7Y7Ywwk-fGazWu-y59xh-hYY15-bzd5ym-5XudVL-cEwUnU-laf3S1-bFQKS6-njRt8-34Wub7-nAyTys-GyMse-3b8udo-hglxR5-5R5ebY-aZk2zz-2WxQ3n-7w9QMBeccf75F-4Fh97Y-3mzwqg-7yhraF-8PFLQb-qA3pC)

Lego: [https://www.flickr.com/photos/defjux/8152400914/in/photolist-kfoWq7-aB6Zno-aB6Zs7-dXFmeW-nyUXb-5YoNCK-hj7ij-paBzB-npBpgo4-baUFu-bPpgUr-baBUY-baUXW-bPpgUH-baAV1-bPpgWB-bPpgLb-bPpgsT-bAucWj-bPpgQF-6PpgYp-bPpgCk-baUJ1-baU90-bPpgHR-5gwDbS-baUcn5-8M2XsG-baUcrA-baBUxU-bxU3Lk-dqbblS-ddm4bP-ddM2qj-ddM4GP-ddM6M2-ddM4Wm-ddM6Xp-ddM2Hp-ddM4kV-ddM3q1-ddM5eV-ddM8YS-ddM8RW-ddM5P0-ddLZRo4-ddM4gR-ddM6my-fGgEFR-7q8DPy/](https://www.flickr.com/photos/defjux/8152400914/in/photolist-kfoWq7-aB6Zno-aB6Zs7-dXFmeW-nyUXb-5YoNCK-hj7ij-paBzB-npBpgo4-baUFu-bPpgUr-baBUY-baUXW-bPpgUH-baAV1-bPpgWB-bPpgLb-bPpgsT-bAucWj-bPpgQF-6PpgYp-bPpgCk-baUJ1-baU90-bPpgHR-5gwDbS-baUcn5-8M2XsG-baUcrA-baBUxU-bxU3Lk-dqbblS-ddm4bP-ddM2qj-ddM4GP-ddM6M2-ddM4Wm-ddM6Xp-ddM2Hp-ddM4kV-ddM3q1-ddM5eV-ddM8YS-ddM8RW-ddM5P0-ddLZRo4-ddM4gR-ddM6my-fGgEFR-7q8DPy/)

Dancers: [https://www.flickr.com/photos/lecercle/845619594/in/photolist-cKFQSE-jorofj-tpG66-694yuv-nn1mM9-2hJ2fb-4YZNm-p9NuYH-5Sm8G5-6524K8-94NaBt-40wWPJ-9h9RZt-4GJ2JL-aPVock-f7VPvF-8TIW7t-aCZKZc-ddqWc6-5xwB8Mm-77K97G-bb5JDB-99rwUF-4HUZoD-63896-cqhbif-cqhbxq-cqhaRU-4HUZvV-2jczhz-5LPrlLJ-aXVac-7tkoY-6z4PwQ-9FA2ZV-2JmKvD-5ZusNy-4q6kaD-8BEJWeW-edfMiF-9taSq6-58gpkGk-jdQLuV-8A7cs7-fDNtzY-ddqW9T-sp4g2-AL6RM-2HvzfL-4Z5FMCC](https://www.flickr.com/photos/lecercle/845619594/in/photolist-cKFQSE-jorofj-tpG66-694yuv-nn1mM9-2hJ2fb-4YZNm-p9NuYH-5Sm8G5-6524K8-94NaBt-40wWPJ-9h9RZt-4GJ2JL-aPVock-f7VPvF-8TIW7t-aCZKZc-ddqWc6-5xwB8Mm-77K97G-bb5JDB-99rwUF-4HUZoD-63896-cqhbif-cqhbxq-cqhaRU-4HUZvV-2jczhz-5LPrlLJ-aXVac-7tkoY-6z4PwQ-9FA2ZV-2JmKvD-5ZusNy-4q6kaD-8BEJWeW-edfMiF-9taSq6-58gpkGk-jdQLuV-8A7cs7-fDNtzY-ddqW9T-sp4g2-AL6RM-2HvzfL-4Z5FMCC)