Using CRIS to power research data discovery

Alex Ball\textsuperscript{1}  Christopher Brown\textsuperscript{2}  Laura Molloy\textsuperscript{3}  
Veerle Van den Eynden\textsuperscript{4}  David Wilson\textsuperscript{3}

\textsuperscript{1}DCC/UKOLN, University of Bath  
\textsuperscript{2}Jisc  
\textsuperscript{3}DCC/HATII, University of Glasgow  
\textsuperscript{4}UK Data Archive

12 May 2015

Except where otherwise stated, this work is licensed under the Creative Commons Attribution 4.0 International licence: http://creativecommons.org/licenses/by/4.0/

Supported by

EuroCRIS Membership Meeting, Paris 2015-05-12
Outline

Vision

Phase 1

RIF-CS

Mapping from CERIF to RIF-CS

Phase 2, with added CRIS?
Scope and purpose

UK Research Data
Discovery Service for UK Research Data
Scope and purpose

Discovery Service for UK Research Data

Data sharing and reuse

- data impact
- funder value
- researcher credit
- better record
Where does RDRDS fit in?

- BADC
- UKSSDC
- ADS
- Oxford
- Edinburgh
- UKDA

NERC DCS

DataCite

Jisc RDDS

International aggregators

Search engines

Other UK registries

data.gov

Search engines

Other UK registries

EuroCRIS Membership Meeting, Paris 2015-05-12
Phase 1 pilot

- Discovery service based on ORCA (Research Data Australia)
- Participating data repositories:
  - UKDA
  - Archaeology Data Service
  - 7 NERC Data Centres: BADC, BODC, EIDC, NEORC, NGDC, PDC, UKSSDC
- Harvested metadata in six standard formats
- Harvested metadata according to three protocols:
  - native XML export
  - OAI-PMH
  - CSW (Catalogue Services for the Web)
  - (but we had to cheat a little bit with the last two)
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.

- **Services** (0)
  Services that support the creation or use of research datasets or collections.

Spotlight on research data

http://rdrds.cloudapp.net/

Who contributes to the Research Data Registry and Discovery Service?

5 research organisations from around the UK contribute information to Research Data Australia.

See All
About RIF-CS

- Profile of ISO 2146 (Information and Documentation – Registry Services for Libraries and Related Organizations)
- Optimized for collection services registries
- Maintained by ANDS: see http://services.ands.org.au/documentation/rifcs/1.6/guidelines/rif-cs.html
- ‘Gateway drug’ for CERIF?
Example set of related objects

**COLLECTION**

- dataset
- CSV file
Example set of related objects
Example set of related objects
Example set of related objects
Example set of related objects

- **PERSON**: J Whiteway
- **PARTY**: NERC
- **SERVICE**: create sensor
- **COLLECTION**: dataset CSV file
- **COLLECTION**: project utls_egrett
- **ACTIVITY**: produces
- **ACTIVITY**: hasOutput
- **ACTIVITY**: hasPart
- **ACTIVITY**: isPrincipal-InvestigatorOf
- **ACTIVITY**: isFunderOf

Repository: BADC

EuroCRIS Membership Meeting, Paris 2015-05-12
Example set of related objects

- **COLLECTION**: BADC
- **PARTY**: NERC
- **ACTIVITY**: UTLS
- **SERVICE**: create sensor
- **COLLECTION**: dataset CSV file
- **PARTY**: person J Whiteway
- **ACTIVITY**: hasOutput
- **ACTIVITY**: hasPart
- **PARTY**: isFunderOf
- **PARTY**: isPrincipal-InvestigatorOf

- **repository**
- **group**
- **program**
- **dataset**
- **project**
- **CSV file**
- **utls_egrett**
- **BADC**
- **NERC**
- **UTLS**
- **person**
- **group**
- **program**
- **J Whiteway**

EuroCRIS Membership Meeting, Paris 2015-05-12
Metadata flows

EuroCRIS Membership Meeting, Paris 2015-05-12
Problem statement
Problem statement

COLLECTION

dataset
  key
  name
  ...

PARTY

group
  key
  name
  ...

PARTY

isPrincipal-InvestigatorOf

ACTIVITY

isPrincipal-InvestigatorOf

isFunderOf

person
  key
  name
  ...

project
  key
  name
  ...

EuroCRIS Membership Meeting, Paris 2015-05-12
From CERIF to RIF-CS: Projects

<table>
<thead>
<tr>
<th>cfProj</th>
<th>ACTIVITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>cfFedId/cfFedId</td>
<td>project</td>
</tr>
<tr>
<td>cfFedId/cfFedId_Class</td>
<td>identifier[@type=?]</td>
</tr>
<tr>
<td>cfProjId</td>
<td>idenfier[@type=local]</td>
</tr>
<tr>
<td>cfProjTitle</td>
<td>name[@type=primary]/namePart</td>
</tr>
<tr>
<td>cfAcro</td>
<td>name[@type=abbreviated]/namePart</td>
</tr>
<tr>
<td>cfURI</td>
<td>location/address/electronic[@type=url]/value</td>
</tr>
<tr>
<td>cfProjKeyw</td>
<td>subject[@type=local]</td>
</tr>
<tr>
<td>cfProjAbstr</td>
<td>description[@type=full]</td>
</tr>
<tr>
<td>cfStartDate*</td>
<td>coverage/temporal/date[@type=dateFrom]</td>
</tr>
<tr>
<td>cfEndDate†</td>
<td>coverage/temporal/date[@type=dateTo]</td>
</tr>
</tbody>
</table>

* deprecated; see also ‘Awarded’ cfProj_Class/cfStartDate, ‘Funder’ cfProj_OrgUnit/cfStartDate.
† deprecated; see also ‘Awarded’ cfProj_Class/cfEndDate, ‘Closed’ cfProj_Class/cfStartDate, ‘Funder’ cfProj_OrgUnit/cfEndDate.
From CERIF to RIF-CS: Organisations

PARTY

group

- identifier[@type=?]
- identifier[@type=local]
- name[@type=primary]/namePart
- name[@type=abbreviated]/namePart
- location/address/electronic[@type=url]/value
- subject[@type=local]
- description[@type=full]
- location/address/electronic[@type=email]/value
- location/address/physical[@type=postalAddress]/addressPart[@type=addressLine]
From CERIF to RIF-CS: Data creators

**PARTY**

**person**

- identifier[@type=?]  
- identifier[@type=local]  
- name[@type=primary]/namePart[@type=family]  
- name[@type=alternative]/namePart[@type=family]  
- name[@type=primary]/namePart[@type=given]  
- name[@type=alternative]/namePart[@type=given]  
- subject[@type=local]  
- location/address/electronic[@type=email]/value  
- location/address/physical[@type=postalAddress]/addressPart[@type=addressLine]
From CERIF to RIF-CS: Datasets

COLLECTION

project

- identifier[@type=‽]
- citationInfo/citationMetadata/identifier[@type=‽]
- identifier[@type=local]
- name[@type=primary]/namePart
- citationInfo/citationMetadata/title
- name[@type=alternative]/namePart
- dates[@type=dc.issued]/date[@type=dateFrom]
- citationInfo/citationMetadata/date[@type=publicationDate]
- location/address/electronic[@type=url]/value
- subject[@type=local]
- description[@type=full]
- coverage/spatial[@type=iso19139dcmiBox]
- citationInfo/citationMetadata/version
- citationInfo/citationMetadata/publisher

From CERIF to RIF-CS: Datasets

- cfProj
  - cfFedId/cfFedId
  - cfFedId/cfFedId_Class
  - cfResProdId
  - cfResProdName
  - cfResProdAltName
  - ???
  - cfURI
  - cfResProdKeyw
  - cfResProdDescr
  - cfResProd_GeoBBox
  - cfResProdVersInfo
  - cfOrgUnit_ResProd
  - ‘Publisher’

EuroCRIS Membership Meeting, Paris

2015-05-12
From CERIF to RIF-CS: Relationships

- 'Principal Investigator' to 'Researcher'
  - cfProj
  - Pers
  - Person Project Engagements

- 'Participant'
  - cfProj
  - Pers
  - Person Project Engagements

- 'Manager'
  - cfProj
  - Pers
  - Person Project Engagements

- 'Funder'
  - cfProj
  - OrgUnit
  - Organisation Project Engagements

ACTIVITY: project

- hasPrincipal-Investigator
  - PARTY: person
  - isPrincipal-InvestigatorOf

- hasParticipant
  - PARTY: person
  - isParticipantIn

- isManagedBy
  - PARTY: person
  - isManagerOf

- isFundedBy
  - PARTY: group
  - isFunderOf

EuroCRIS Membership Meeting, Paris
2015-05-12
From CERIF to RIF-CS: Relationships

- Project Output Roles:
  - cfProj
  - ResProd
  - ‘Originator’

- Person Organisation Roles:
  - cfPers
  - OrgUnit
  - ‘Affiliation’

- Person Output Contributions:
  - cfPers
  - ResProd
  - ‘Creator’?

- Inter-Product Relations:
  - cfResProd
  - ResProd
  - ‘Part’

- Activity Collections:
  - ACTIVITY
  - project
  - hasOutput
  - isOutputOf
  - COLLECTION
  - dataset

- Person Collections:
  - PARTY
  - person
  - isMemberOf
  - hasMember
  - PARTY
  - group

- Contribution Collections:
  - PARTY
  - person
  - isPrincipal-InvestigatorOf
  - hasPrincipal-Investigator
  - COLLECTION
  - dataset

- Collection Relations:
  - COLLECTION
  - collection
  - hasPart
  - isPartOf
  - COLLECTION
  - dataset

EuroCRIS Membership Meeting, Paris 2015-05-12
RIF-CS could also express

- the party that manages the dataset;
- the party that owns the dataset;
- a publication that cites the dataset;
- a publication that documents the dataset;
- a publication to which the dataset is a supplement.
Example party record: without CRIS

Alan Cresswell

Connections

Collections

Validated radiometric mapping in 2012 of areas in Japan affected by the Fukushima-Daiichi nuclear accident
Example party record: with CRIS?

CRAVEN, MIKE
Australian Antarctic Data Centre

Contact Information
m.craven@utas.edu.au
Ph: +61 3 6226 7876
Fax: +61 3 6226 6902
CRAVEN, MIKE
203 Channel Highway
Australian Antarctic Division
Tasmania 7050
Australia

Related Data
- Processor Ice shelf - ocean interaction in the cavity beneath the Amery Ice Shelf
- Processor Amery Ice Shelf - hot water drill borehole, 2001-02 - AM01
- Processor Amery Ice Shelf - hot water drill borehole, 2001-02 - AM01 Aquadopp current meter data
- Processor Amery Ice Shelf - hot water drill borehole, AM01 Brancker thermistor data
- Processor Amery Ice Shelf - hot water drill borehole, 2001-02 - AM01 FSI-CTD profile data

View all 52 related data

User Contributed Tags

Login to tag this record with meaningful keywords to make it easier to discover

EuroCRIS Membership Meeting, Paris
2015-05-12
Merci beaucoup pour votre attention
Thank you for your attention

DCC Website: http://www.dcc.ac.uk/

UK RDDS Phase 2 home page: http://www.jisc.ac.uk/rd/projects/uk-research-data-discovery