Phase 1 RDRDS Metadata

Alex Ball

DCC/UKOLN, University of Bath

16 June 2014
RDRDS Metadata

RIF-CS
We are not committed to RIF-CS
We are not committed to RIF-CS

(but it works quite well)
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/1.5/guidelines/rif-cs.html
- ‘Gateway drug’ for CERIF?
Data model

- COLLECTION: repository
- PARTY: group
- ACTIVITY: program
- SERVICE: create

- COLLECTION: collection
- PARTY: person
- ACTIVITY: project
- SERVICE: generate

- COLLECTION: dataset
- PARTY: administrative-Position
- ACTIVITY: course
- SERVICE: transform
Example set of related objects

COLLECTION

- dataset
- CSV file
Example set of related objects

SERVICE
create
sensor

produces

collection

dataset

CSV file

ACTIVITY

hasOutput

program

UTLS

ACTIVITY

hasPart

group

NERC

PARTY

isFunderOf

person

J Whiteway

PARTY

isPrincipalInvestigatorOf

isPrincipalInvestigatorOf

repository

BADC

COLLECTION

isLocationFor

Jisc RDRDS Phase 2 Kickoff, London 2014-06-16 #jiscrdrds
Example set of related objects
Example set of related objects:

- **create sensor**
- **dataset CSV file** produces
- **group NERC**
  - **PARTY**
  - **isFunderOf** **program UTLS**
  - **hasPart** **ACTIVITY**
    - **isPrincipalInvestigatorOf**
      - **isFunderOf**
        - **PARTY**
          - **isFunderOf**
            - **COLLECTION**
              - **produce**
                - **ACTIVITY**
                  - **hasOutput**
                    - **repository BADC**
                      - **isLocationFor**
                        - **ACTIVITY**
                          - **hasPart**
                            - **PERSON J Whiteway**
                              - **PARTY**
                                - **isFunderOf**
                                  - **ACTIVITY**
                                    - **hasPart**
                                      - **ACTIVITY**
                                          - **hasOutput**
                                            - **program utls_egrett**

Example set of related objects

- **GROUP**
  - NERC

- **PROGRAM**
  - UTLS

- **SERVICE**
  - create sensor

- **COLLECTION**
  - dataset CSV file

- **ACTIVITY**
  - hasOutput
    - utls_egrett
  - hasPart

- **PARTY**
  - isFunderOf
    - UTLS
  - isPrincipal-InvestigatorOf
    - J Whiteway

- **PARTY**
  - isFunderOf
    - NERC
  - isPrincipal-InvestigatorOf
    - J Whiteway

**Repository**
- BADC
  - isLocationFor

**Event**
- Jisc RDRDS Phase 2 Kickoff, London 2014-06-16 #jiscrdrds
Example set of related objects

- Collection: repository BADC
- Service: create sensor
- Party: group NERC
- Collection: dataset CSV file
- Activity: program UTLS
- Activity: project utls_egrett
- Party: person J Whiteway
- Party is funder of program UTLS
- Activity has output project utls_egrett
- Collection is location for repository BADC
- Service produces dataset CSV file
- Party is principal investigator of person J Whiteway
We are not committed to RIF-CS
Metadata crosswalks

DataCite 3
- Archaeology Data Service
- Oxford

DDI Codebook 2.5
- UK Data Archive

EPrints 3
- Glasgow
- Leeds
- Southampton

MODS 3.5
- Edinburgh
- St Andrews
- Hull

OAI-PMH Dublin Core
- Oxford Brookes
- Lincoln

UK Gemini 2.2
- NERC Data Catalogue Service (incl. ADS)
Artist’s impression of a crosswalk

- PARTY
  - group
  - originatingSource
  - name

- COLLECTION
  - group
  - originatingSource
  - key
  - name
  - description[full]

- OAI_DC
  - request
    - header
      - identifier
    - metadata
      - dc:creator
      - dc:title
      - dc:description
      - ...

hasPrincipal-Investigator

isPrincipal-InvestigatorOf

Jisc RDRDS Phase 2 Kickoff, London 2014-06-16 #jiscrdrds
HTTP or OAI-PMH

**HTTP**
- DataciteToRifcs (Single XML record)
- EprintsToRifcs (EPrints XML export)
- Gemini2p2ToRifcs (CSW)

**OAI-PMH**
- Datacite3ToRifcs
- Ddi2p5ToRifcs
- Mods3ToRifcs
- OaiDcToRifcs
Quality levels for collections

Quality Level 2

- title
- description
- location (e.g. URL)
- IPR statement
- related party, e.g.
  - P.I./researcher
  - manager

Quality Level 3

- identifier
- citation information*
- subject
- date (e.g. of publication)
- spatial coverage
- temporal coverage
- related activity

* Such as ‘publisher’; other relevant fields are already mentioned.
Experiences

- RIF-CS can handle
  - ‘stub’ records with minimal information;
  - structured information in structured way;
  - unstructured information in unstructured way.
- We needed to expand the controlled vocabulary for subject schemes.
- RIF-CS does not describe what web links do.
- Parties need IDs too.
- There’s no specific, direct relation between a funder and a dataset (it goes via the grant).
Still to do...

- Harvesting a new version of a record replaces the old one.
  - How do we merge into the old one?
  - How do we conditionally replace the old one?
  - How do we handle deletions?
- Which dates do we really need?
- How do we get ‘boilerplate’ information from user accounts?
- How do we harvest from CRISes in CERIF format?
We are not committed to

RIF-CS
We are not committed to RIF-CS

Would something else work better?
Thank you for your attention

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