US Rolling Deck to Repository (R2R) Program

Data Discovery and Access
Open Access

Initial focus on *High-level discovery + Machine-level interfaces.*

R2R content published online:

1. Cruise summaries as ISO 19115 standard metadata records
2. Cruise tracks (control point geometries) in OGC Web Feature Service (WFS)
3. Vocabularies +mappings as SKOS collection
4. Entire metadata catalog (Cruises, Vessels, Sensors, Datasets, Persons, etc) as Linked Data collection
5. Datasets registered with DOIs + standard metadata records
1. Cruise Summaries

ISO template (profile-less) developed jointly with US NOAA National Data Centers

Add URIs for controlled terms in ISO records eg.

```xml
<gmd:MD_Identifier>
  <gmd:code>
    <gco:CharacterString>KM0715</gco:CharacterString>
  </gmd:code>
</gmd:MD_Identifier>

<gmd:MD_Identifier>
  <gmd:code>
  </gmd:code>
</gmd:MD_Identifier>
```
2. Cruise Tracks

OGC Web Feature Service (WFS) – returns minimal control points as GML LINESTRING geometry

http://get.rvdata.us/wfs?service=wfs&version=1.0.0&request=getcapabilities
3. SKOS Vocabularies

Concept Collection published as hierarchy of terms eg.

http://voc.rvdata.us/

http://voc.rvdata.us/device

http://voc.rvdata.us/device/gravimeter

R2R term mapped to NVS term
4. Linked Data

Why? Linked Data approach satisfies requirement to make R2R Catalog content broadly and easily accessible:

- Publish authoritative resources (Cruises, Vessels, Sensors, Datasets, Persons, etc) with well-structured URIs that can be dereferenced
- Combine with standard query interface (SPARQL)
- Engage partner data systems pursuing similar approach with closely-related data
Linked Data: Partners

- Biological & Chemical Oceanography Data Management Office (BCO-DMO)
- Index to Marine & Lacustrine Geological Samples
- InterRidge Hydrothermal Vents Database
- Integrated Ocean Drilling Program – U.S. Implementing Organization (IODP-USIO)
- NERC Vocabulary Server
- Rolling Deck to Repository (R2R) Program
Linked Data: Implementation

Deployed D2R Server (http://d2rq.org/d2r-server/) that maps existing SQL database into RDF on the fly.

Provides

- (backend) SPARQL query endpoint for machines
- (frontend) Pubby browser for humans

So far, performance is good for ~tens of thousands of triples.

http://linked.rvdata.us/
Linked Data: Example

1. Classify R2R instruments against NVS vocabulary

2. Map all datasets produced by instrument

3. Discover relationships among instruments
Linked Data: Best Practices

Heath and Bizer (2011), Linked Data: Evolving the Web into a Global Data Space

• 303-style (not Hash-style) URIs
• map to resources, ideally external, not literals eg. `dcterms:lang <http://id.loc.gov/vocabulary/iso639-2/eng>` instead of `dcterms:lang "eng"`
• `rdfs:label` property for every resource
• `dcterms:source` link to dataset (+ :tableOfContents)
• `geo:lat` and `geo:long` for position
Linked Data: Lessons Learned

• Forced us to clearly differentiate concepts (vocabularies) from instances

• Forced us to move knowledge+relationships *out* of the search interface and *into* the data resources

• Forced us to start talking to partner systems and aligning vocabularies
5. Data Citation

• R2R will publish Digital Object Identifier (DOI) for each dataset accessioned to a long-term archive

• Contract with US “EZID” system, to register DataCite DOIs

• Follow best practices developed by Earth Science Information Partners (ESIP) Federation

http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/Principles
Data Citation Example

CTD data from Cruise OC404-01 (EDDIES, 2004 Survey 1)

1. Original Data (R2R) doi:10.7284/X

2. Final Data doi:10.1575/1912/4847

Next Steps

- Map Cruise tracks to Global Gazetteers eg. EEZs, UFNs
- Map Datasets to Parameters (P02)
- Specific ODIP Deliverables per R2R Funding Request
  1. Develop or Crosswalk additional vocabularies
     - Instruments (L22), Units (P06), Ports (C38)
     - Event activities
     - Quality Control flags – MET data
  2. SPARQL 1.1 Federated Discovery
  3. EU Standards Integration
     - Crosswalk Cruise metadata to SeaDataNet CSR ISO Profile
     - Deploy WMS for Cruise tracks as INSPIRE View Service