Collaborative Information Management in Scientific Research Processes

Shirley Crompton, Brian Matthews, Erica Yang: STFC Scientific Computing
Cameron Neylon: STFC ISIS
Simon Coles: University of Southampton, School of Chemistry
Science and Technology Facilities Council

- **Research Funder**
  - Provides large-scale scientific facilities for UK Science
    - ISIS, Diamond, CLF…

- **Scientific Computing**
  - Facilities e-Infrastructure
    - ICAT Data Catalogues
    - E-Publications Archive
    - Petabyte Data Store
  - Active contributor to data management, digital data preservation research and infrastructure building
    - ICAT, SRF, Scidip-ES, SCAPE …
    - PanDATA, APARSEN…
Overview

• The Webtracks project
• Our motivations
• The Webtracks approach
• Webtracks exemplars
• Future Work : SRF
• Summary
Webtracks Project

- Web-scale link TRACKing for research data and publications
- Develop an approach and mechanism for constructing and propagating linked data in the context of research activities
- Building on prior work in Claddier, StoreLink on citation linking
- Collaboration between:
  - Chemistry Dept, Southampton University
  - STFC Scientific Computing
- Funded by the JISC Managing Research Data (MRD) Programme

http://www.jisc.ac.uk/whatwedo/programmes/mrd/clip/webtracks.aspx
The Scientific Research Process (1/2)

Diverse tools/services + Objects

- Experiment
- Sample
- Raw data
- ICAT
- Data Catalogue
- Data analysis
- Derived data
- Describe derived data
- Electronic lab notebook
- Describe sample
- Describe algorithm
- Protein model
- Publication
The Scientific Research Process (2/2)
Contents Dispersed

- Experiment
- Raw data

STFC
Data Catalogue

Sample

Describe sample

Electronic lab notebook

Data analysis

Derived data

Describe derived data

STM

Publication

Protein Data Bank

On-line Journal

Publication

Institution

Home

Describe algorithm
Adding Structured Citation Links

- Experiment
- Raw data
- Sample
- Data Catalogue
- Electronic lab notebook
- Creator
- Has PI
- Provides Data For
- Provides Method For
- Has Sample
- Describe sample
- Describe algorithm
- Describe derived data
- Derived data
- Documents is Documented By
- Is Cited As DataSource
- Publication
- Publication cites is Cited By
A Federated Web of Research Objects

A Grand Experiment

cito: cites

cito: provides

dc:title
If we allow repositories to communicate ....
If we allow repositories to communicate ....
Querying the RDF graph to form Research containers
Research containers

Fragments of Annotated RDF graph capturing specific relationship between on-line research objects

Example usages....

- My Project
  - Provenance tracing

- My Paper
  - Discovery of related objects

- My Citations
  - Impact factor

- My Biblio Search
  - Research resource

- My Expr
  - Trusted publication
The Webtracks Approach

- Inter-repository communication protocol
  - Based on Storelink

- Webtracks Java Restlet application
Related protocols

• Trackbacks
  • track distributed blog ‘conversations’
  • With fixed semantics
• Semantic Pingback
  • RPC protocol
• Salmons
  • RSS protocol, non-RDF approach
• Claddier/Storelinks
  • P2P, no centralised service
  • Support fixed formats of metadata and link semantics
• ....
The Webtracks InteRCom protocol

Repository A (Link Publisher)

Resource Layer

Resource A

Res B metadata

link property

HTTP Layer

HTTP requests:

GET metadata

POST link

Repository B (Link Receiver)

Resource B

Sender

Receiver

A Grand Experiment

R1 (Perma link)

cito: cites

cito: isCitedBy

R2 (DOI)

dc:title

UNIVERSITY OF SOUTHAMPTON

Science & Technology Facilities Council
e-Science

ESC2012
Architecture (main classes)
(based on Restlet Framework)

ObjectsFacade

DataFacade

MyFacade

Data Layer Access

Application

IntercomApp

MyApp

App-Spec config (eg Security, Encoding, Tunnelling …)

ServerResource

Resource

MyResource

Resource Wrapper

Extension point

Restlet 2.1
Webtracks
User App
Webtracks Resource Info Model

- Repository
  - A Repository
  - Resource
    - A Resource (digital research object)
      - Link
        - A Link associated with the Resource
          - Each link has a subject, predicate and object
            - Subject
            - Predicate
            - Object
Webtracks-Icat Exemplar

- Investigation Resource
  - DOI Landing Page
  - HTML rep with RDFa

- RDF rep

- Links Resource
  - HTML rep with RDFa
Webtracks-Epubs Exemplar

Epubs

Expression Resource
- HTML rep with RDFa

RDF rep

Links Resource
- HTML rep with RDFa
Smart Research Framework - experiments and the research life cycle

Proposal
Scientist submits application for beamtime

Approval
Facility committee approves application

Scheduling
Facility registers, trains, and schedules scientist’s visit

Experiment
Scientists visits, facility run’s experiment

Metadata Repository

Data cleansing
Raw data filtered and cleansed

Data analysis
Tools for processing made available

Publication
Subsequent publication registered with facility

WebTracks

Smart Research Framework
- in process provenance capture

Metadata Repository

Proposal
Approval
Scheduling

Experiment

Data Storage & cataloguing
Publication
Data analysis

Sample Registration
Experiment Setup
Data Tracking & Collation
Data Reduction

(Orange boxes: processes targeted in the SRF project)
Smart Research Framework
- Schematic Scenario

ELN

Instrument Control Script

Beamline

SRF Engine

Mantid
(Reduction)

Beamsight
(Model Fitting)

Blog Posts
in ELN

Data Catalogue

ICAT

File Store

EPubs

Linking With Webtracks
Smart Research Framework
- blog posts capturing lifecycle events in an experiment
Summary

Collaborative Information Management via the intelligent pushing of citation links

• Leverage emerging Linked Data environment
• Link diverse types of digital research objects
• Facilitate access and discovery of related objects
• Restore context to dispersed digital research outputs
• No constraints on link semantics and metadata
• P2P, does not rely on a centralised service
• Flexible usage, link can be propagated at any time
• Mix and match research object networks from aggregated citation links according to the link resource/s or property/ies of interest.

• ....
Webtracks:
http://www.jisc.ac.uk/whatwedo/programmes/mrd/clip/webtracks.aspx
SRF : http://www.mylabnotebook.ac.uk/
Webtracks Download:
http://sourceforge.net/projects/webtracks/files/