The Global Registries Initiative: Progress Report and Software Demonstration

Chris Blackall, Australian National Data Service
Jeremy Frumkin, University of Arizona
Vic Lyte, University of Manchester
Who

• Jeremy Frumkin, Assistant Dean/Chief Technology Strategist, University of Arizona Libraries
  – Ockham Initiative http://www.ockham.org/

• Vic Lyte, Project Director, University of Manchester
  – Information Environment Service Registry http://iesr.ac.uk/

• Chris Blackall, Research Analyst, Australian National Data Service (ANDS)
  – ANDS Collections Registry https://services.ands.org.au/home/orca
Welcome to the Global Registries Initiative

The global registries initiative aims to make scholarly information more discoverable by, and useful to, the research community and provide a shared framework to exploit the opportunities offered by Web 2.0 technologies. A coordinated international mesh of collection/service registries will enable better discovery and innovative re-use of scholarly collections.

Collections/service registries record information about collections of scholarly materials and the web services available to exploit this content in powerful ways.

When scaled internationally, it is anticipated that a network of collections and services registries would be beneficial.
What

- http://globalregistries.org
- Collection/service registries record information about ‘collections’ of scholarly resources and related web services
- Collections? ‘Any grouping of digital items’
- Global? ‘register locally, discover globally’
- It works…stick around for the demo
Why

• Managing the data deluge
• Supporting data-intensive science (e-science)
• Federating of research data collections
• Some questions:
  – Why not leave it to Google?
  – Why not use Semantic Web/Linked Data
    http://linkeddata.org/
How

• Collections and services records are harvested locally and aggregated nationally/globally
  – ‘Federated’ registry architecture
  – Metadata harvest protocol
  – Registry interchange formats

• Policies and metadata sharing agreements
Identifier: http://registry.library.oregonstate.edu/collection/871
Associated Agent: National Science Digital Library
Available Services:
arXiv.org E-Print Archive

Subject Keywords:
Physics; Astrophysics; Condensed Matter; Disordered Systems and Neural Networks; Materials Science; Mesoscopic Systems and Quantum Hall Effect; Soft Condensed Matter; Statistical Mechanics; Strongly Correlated Electrons; Superconductivity; General Relativ
Computer science--Collections
Physics--Collections
Mathematics--Collections
Physics
Mathematics
Applied mathematics
Science
Computer science
Engineering, Computing, and Technology
Mathematics
Physical Sciences

Description:

ArXiv.org, the major e-Print archive in the areas of physics, computer science, mathematics and related disciplines, makes available online research articles in relevant disciplines prior to print publication. Articles are submitted by their authors, and are then available freely, in various formats.

Collection Size:
Collection Type: Text
Logo URL: http://nsdl.org/images/brands/5101.gif
Rights:
Use Rights:
Access Rights:
Item Format: text
The Information Environment Service Registry (IESR) is the UK’s free catalogue of electronic resources;

Sponsored by the Joint Information Systems Committee (JISC) and part of shared services layer in overall IE landscape;

The IESR aims to provide a reliable source of descriptive information about electronic resources and technical details about how they can be accessed via web or machine interfaces;

http://iesr.ac.uk
IESR – Main Challenges

• Initial concept too technically abstract – had to have a way to engage non-technical stakeholder community – What are the benefits;
• High access barrier to easy and ubiquitous use – detailed metadata schema required too much effort for collection and resource contributors;
• Restricting underlying technology – needed to be more scalable, adaptive and ubiquitous;
• Wide range of collection / resource / artifact types potentially in scope;
• Need to optimize sister and allied search and discovery services:
  – www.intute.ac.uk/irs/demonstrator
• Need end-to-end clear (non-technical) use cases.
• Need stakeholder and organisational involvement / steerage / buy-in – What are the benefits and how does it improve end-user experience?
IESR – Current priorities & activities

- Development of an outreach service and re-branding – how can we add benefit to you;
- Developing a minimal metadata description set, based on user requirements, supported by in-house cataloguing resource;
- Embedding advanced technologies ‘under the hood’ with open interface on top – Autonomy IDOL & Lucene for automatic metadata generation, dynamic discovery (clustering, concept matching) and unstructured search;
- Establishing early and visible wins:
  - Search engine optimization;
  - National Health Service;
  - Regional / National life-long learning initiatives;
  - Cross-sector knowledge initiatives;
  - Academic-oriented ‘Yellow-Pages’ for electronic resources.
- Academics want Global search scope ⇒ need for GRI topology to expose related assets to search services
Australia

• Collection/Service Registry initiative is a central part of national research infrastructure funded by the Australian Government

• Focus on building a Research Data Commons: data discovery and citation services

• Drivers: Innovation and research evaluation
ISO 2146 Registry Services for Libraries and Related Organizations
Collection

Type: collection
Title: Community-based catch-monitoring of dugongs (Dugong dugon) and green turtles (Chelonia mydas) in the Kaiwalagal region of Torres Strait 2006

Description: This study trialled community-based catch-monitoring strategies for dugongs and green turtles in two communities (Hammond and Thursday Islands) in the Kaiwalagal region (inner islands) of Torres Strait. Hunters completed datasheets after each hunting trip and these were collected weekly by Indigenous catch-monitors employed on the project. The catch-monitoring program was evaluated in terms of the reliability of the information collected and the suitability of the method to the communities. The information collected included biological information about the composition of the catch (e.g. sex ratio, size distribution, reproductive status), selectivity of the green turtle catch, catch-effort, catch-gear and the uses for which dugongs and turtles were hunted. The participating communities can use the information in future management of dugongs and green turtles in their communities.

Address: http://www.marine.csiro.au

Subject: 050205 (anzsrc-for) Environmental Management
050201 (anzsrc-for) Aboriginal and Torres Strait Islander Environmental Knowledge
050202 (anzsrc-for) Conservation and Biodiversity
060205 (anzsrc-for) Marine and Estuarine Ecology (incl. Marine Ichthyology)

Custodian: Grayson · Jillian
James Cook University
Next Steps

• Formation of working groups
  – Technical
  – Strategic / Organizational / Policy

• Set of experiments testing GRI
Demonstrator 1: ANDS Pilot

• The ANDS Collection/Service Registry global pilot: This demonstrator aggregates records from three collection/service registries into a combined registry run by the Australian National Data Service (ANDS)
• The aggregated pool of records is from Australia, UK, and USA, and can be searched and browsed
• Aggregation is performed using the OAI-PMH (Protocol for Metadata Harvesting) with the RIF-CS Interchange format
• https://devel.apsr.edu.au/cosi/orca/search.php
Demonstrator 2: LibraryFind

• This demonstrates the usage of meta-search, or the querying of registries distributed over three different continents
• It supports distributed search protocols such as z39.50, SRU/SRW, and Open Search, as well as OAI-PMH aggregation
• http://rs.library.oregonstate.edu