

The Global Registries Initiative: Progress Report and Software Demonstration

Chris Blackall, Research Analyst,
Australian National Data Service (ANDS),
Australian National University

Jeremy Frumkin, Assistant Dean/Chief Technology Strategist,
University of Arizona Libraries

Vic Lyte, Project Director, University of Manchester

Over the last two years, key stakeholders in the USA, UK, and Australia have convened a series of meetings to address the need for a global network of digital library collection and service registries to facilitate information discovery and to orchestrate many collection management and access tasks through shared 'services' (<http://globalregistries.org/meetings.html>).

These meetings brought together diverse communities to explore what steps would need to be taken to link registry technologies together, in an interoperable fashion, with the existing repository networks and the growing number of mass-data storage systems that manage and archive research data.

Although the architecture and standards to be used for the global network of registries have yet to be finalized, we are progressively gaining a clearer picture of an optimal architecture and have developed software that demonstrates its potential. Indeed, we showcase and discuss two such software systems:

- (1) A combined collections and services registry run by the Australian National Data Service. This software aggregates metadata records harvested from repositories in USA, UK and Australia for discovery and service orchestration (<https://devel.apsr.edu.au/cosi/orca/search.php>). The metadata aggregation is performed using OAI-PMH (the Open Archives Initiative Protocol for Metadata Harvesting) in conjunction with a registry interchange format (<http://globalregistries.org/rifcs.html>) based on the draft ISO2146 (Registry Services for Libraries and Related Organizations) standard.
- (2) The *LibraryFind* Global Pilot demonstrates the querying of collections and services registries distributed over three continents. It supports distributed search (or meta-search) protocols such as z39.50, SRU/SRW, and Open Search, as well as OAI-PMH aggregation (<http://apollo.library.oregonstate.edu:3001/record/search>).

Both of these systems demonstrate the benefits of a global registry infrastructure. Using these systems we have been able to improve the discoverability of research datasets that would be otherwise part of the invisible 'dark web'. Moreover, they can easily expose collection metadata to the Google crawler and thus datasets etc. appear in Google search results and often rank highly. Realizing the benefits of the software services available through these registries will take more time and requires parallel developments in the repository development domain.

Following the software demonstration we provide a progress report of the Global Registry Initiative and a brief overview how and why we think it will benefit the open repositories community.

In summary, we accept as a given that institutional repositories (as represented by DSpace, Fedora and ePrints repository platforms to name a few) have successfully provided many organizations with a stable technology to archive digital scholarly resources and provide community access to them. The fact that many organizations find it difficult to recruit new content into their repositories is a separate, although not unrelated, issue.

Despite the relative success of institutional repositories, we are increasingly aware of the limitations of the original monolithic 'one-size-fits-all' model. And despite the fact that the main repository development groups that are constantly reengineering their software, we are still struggling to make

progress. To be fair, these limitations are not an issue for organizations where a standalone repository more-or-less addresses their needs; however, we work for agencies that are respectively building national information infrastructures, also referred to as cyberinfrastructure. This involves integrating repositories, data management systems and high-performance computing facilities into an integrated set of systems and services closely aligned to the information needs of the research community.

Based on our assessments, the original 'institutional' repository platforms fundamentally lack the 'cyberinfrastructure-ready' level of integration we require. We identify two areas where immediate improvements can be made. Firstly, repositories need to consistently support collection-level metadata so that research datasets and other complex 'works' can be better described and made discoverable. Secondly, repositories need to lower the costs of creating new metadata for producers and provide mechanisms to improve its quality over time. Lowering the costs of creating metadata can be achieved through tools, processes and workflows that, where possible, automate metadata creation. Improving metadata quality depends upon populating metadata records with descriptive and administrative information about key entities (researcher names, organization names, subject vocabularies etc.). Ideally, this type of information should be sourced from registries. The general principle at work here is that 'authority' information can be produced and consumed by any repository and third-party software, but it is best stored and maintained separately in national and international registries.

Establishing such registries will be neither simple nor cost-free. It requires international agreements and standards and common mechanisms for metadata to be reliably exchanged between registries (and repositories). Hence, we are committed to establishing common architectures and policies for registries that will operate within a 'federated' governance framework.

To illustrate how this framework might work, we discuss our development of small network 'collection and service' registries and how we successfully aggregated of metadata records from each using OAI-PMH. For instance, we discuss how the Ockham registry (<http://registry.ockham.org>), while created originally for the NSDL (National Science Digital Library) community, can now also be used by libraries and institutions outside of the NSDL to discover its valuable collections and services. In addition to Ockham, we briefly present the work of similar registry efforts in the UK, Australia, and Europe.

Overall, our presentation will be of particular interest to repository developers and managers who are interested in providing end-users with new mechanisms for discovering and accessing collections, particularly in the context of the emerging cyberinfrastructure.