

Register Locally – Discover Globally

Report of the Global Registries Meeting, Washington DC, 2007

This is a report of a meeting held in Washington DC on December 9, 2007, in conjunction with the CNI meeting and 3rd DCC Conference, to discuss the establishment of a global network of interoperable registries.

As currently envisaged, this network would exchange metadata about digital *collections* stored in repositories and data centers, and their associated Web services, to provide innovative information services to the public.

Jeremy Frumkin, Adrian Burton and Ann Apps organized the meeting. Each leads a registry project: OCKHAM Initiative¹ in the USA, ORCA-Registry² in Australia, and the Information Environment Service Registry³ (IESR) in the UK respectively.

In attendance were representatives from a wide range of data collecting organizations from around the world with active interests in collections and services registries (see attendee list).

Jeremy Frumkin opened the meeting with introductions around the table and provided the overall context for the meeting. He explained how registries for digital collections and services have emerged to meet different needs, such as for better collection description, discovery and access in relation to the holdings of institutional repositories and mass data storage facilities.

Jeremy pointed that the OCKHAM Initiative, ORCA-Registry and IESR are already exchanging ideas about how to share collections and services metadata through the adoption of common metadata standards and Web Service protocols and interfaces, with a view to discovering research collections internationally and across disciplines. He also argued that the timing was right to broaden this discussion and listed other registry initiatives in the USA that could contribute to the discussion; most notably, the American Social Life Online⁴ as part of the DLF Aquifer project.

Furthermore, he argued that the ability to scale such a network internationally would be advantageous for research communities who need increasingly to manipulate research materials from sources beyond their own community or domain. Indeed, he suggested that the strategic policy development in the cyberinfrastructure domain indicated that the time has come for a co-coordinated approach to registry development. Such development, he maintained, would benefit the users of large datasets and other scholarly curated databases, online collections, and so on—many of which are still largely invisible to most generic Web search engines.

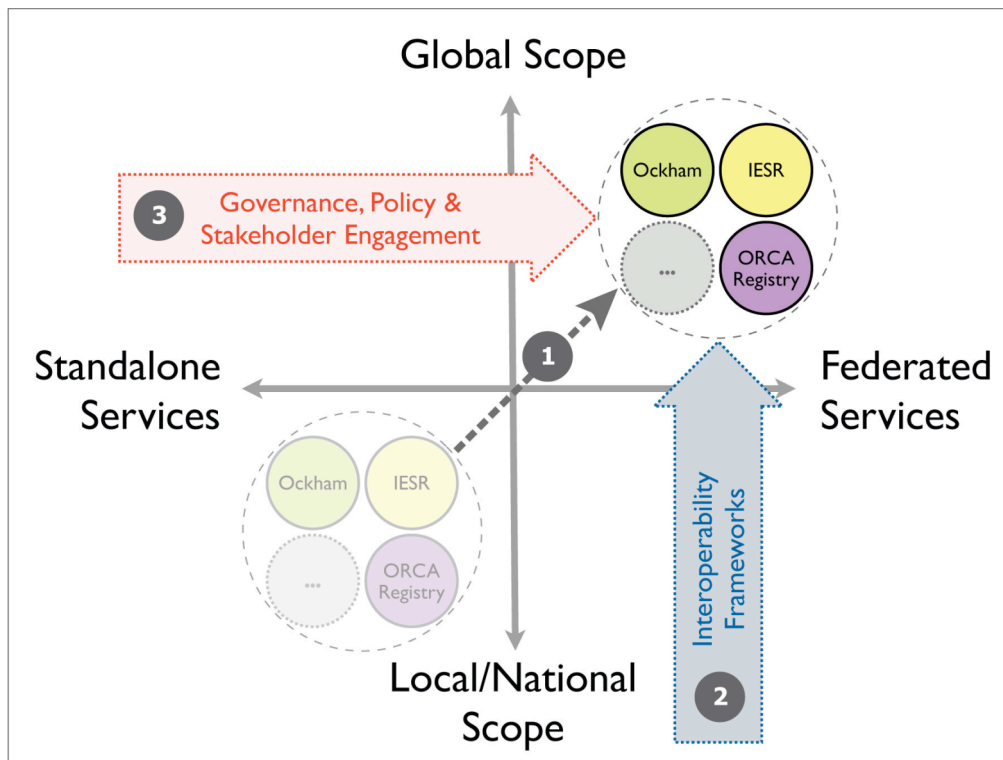
Underlying the importance of this point for the USA-based attendees was the National Science Foundation's Office of Cyberinfrastructure (OCI) calls for proposals for the Sustainable Digital Data Preservation and Access Network Partners (DataNet) initiative⁵. Lucy Nowell from the OCI attended the meeting and expressed interest in how the global registry initiative might benefit the development of cyberinfrastructure in the USA and beyond.

Following Jeremy, Adrian Burton outlined the meeting agenda and identified two key themes to be addressed. These included (a) the desirability, or value

proposition, to stakeholders of a global network of interoperable registries; and (b) the mechanisms for carrying it out (cooperation between international, national and discipline-specific registries etc.).

As part of leading the discussion about the desirability of a global network of interoperable registries, he clarified the proposal’s central concept, summarizing it by referring to a presentation slide (see fig. 1).

Figure 1 Registry scope and configuration model



Drawing attention to the slide, Adrian explained that the current registry solutions are constrained by fact that they have limited (local or national) scope and are typically implemented or configured as standalone applications, with little or no interoperability with other information systems. Thus the goal should be to transform these stand-alone registry instances into a network of federated registries that have global reach (see label 1 on fig 1).

He rejected the idea of constructing a global ‘super registry’, which he maintained would be both technically unfeasible and economically unsustainable. Rather, it could function by stakeholder groups working together collaboratively to develop and adopt of common registry standards and protocols, and by sharing software code, expertise and so on were relevant.

Adrian coined the motto, “register locally—discover globally”, to reflect how a global collections and discovery service might emerge from the sum of its parts.

Finally, Adrian suggested that, for a federated network of registries to be established, simultaneous progress would be required on at least two fronts: interoperability frameworks to enable federated registry systems (label 2); and governance, policy and stakeholder engagement to ensure that provided a consistent quality of service (label 3).

Following this overview, Adrian led the discussion over three sessions. The first focused on the general feasibility of a global network or interoperable registries, the second on the interoperability and governance challenges that would need to be addressed, and the third to discuss the strategy for moving the proposal forward.

In the first session⁶ Adrian raised the desirability of a global network of interoperable registries and then the broader issue of its feasibility. In response, no one raised any fundamental objections to the concept; however, readers should not interpret this as agreement!

As for its overall feasibility, the attendees raised a number of challenges that they thought must be addressed before substantive progress could be made. The challenges were not only confined to registries, but to underlying trends in repository development also.

- It was agreed that are significant human challenges involved in data management on the scale proposed. For example, ensuring that research data is deposited in the first place is a major challenge, let alone registering with global service. Some thought that there was a role here for funding bodies to insist on data deposit in authorized repositories and data centers.
- Researchers will be increasingly required to deposit their data in repositories and data centers in order to ensure certification, access and so on. Funding is likely to become contingent on this; hence, collection registries would be needed.
- Registry developers needed to address rights management and intellectual property issues, especially relating to legacy collections: but how to do this globally?
- Questions were raised about the reach of the proposal. It was pointed out that the meeting did not include representatives from Asia, Africa, Central & South America etc. nor the European Union as such. It was agreed that broadening a set of stakeholders should be considered, as well as multilingual support by registries.
- Was the focus on research data collections inclusive or exclusive of teaching and learning resources? It was pointed out, for example, that IESR contains both, and there was no reason why teaching and learning collections should be excluded. It was suggested that co-operation of e-learning standards groups, such as IMS Global Learning Consortium⁷, would be required.
- Again on the scope of collections to be registered, there were suggestions that many other online collections might need to be considered for inclusion, e.g. Flickr.
- Similarly, the relationship between collection registries and new models for collection description arising from the RDA/FRBR based bibliographic initiatives was discussed.
- The relationship between data collections and data processing was discussed. It was suggested that the size of data sets might be such that computation needs to be done at point of origin. This is being discussed in

the context of DataNet initiative where some collections are of peta-scale or exo-scale.

- It was agreed that as research collections becomes more visible, they would have to be more secure. It was pointed out that data security issues are being dealt with at a national level by designated organizations, so it would be preferable to leave work on this to them. How security would work in a federated network of registries is still an open question.
- Questions were raised about the quality of metadata available from a global network of registries. How would we know if it came from trusted sources, or that it was accurate, up-to-date and comprehensive?

The second session focused on interoperability frameworks and governance challenges.

The main talking point was the choice of registry metadata standards. The acceptance of the *Dublin Core Collections Applications Profile* was noted and discussed, as was the progress being made with the ISO2146 (Registry Services for Libraries and Related Organizations)⁸. The attendees from the Australian Partnership for Sustainable Repositories discussed their positive experiences using the draft ISO2146 to guide the development of the pilot ORCA-Registry service⁹.

Other talking points included how to expose collections and services metadata via OAI-PMH in standardized ways, and the need to develop a metadata interchange format for registries that would enable the consistent harvesting, synchronization and replication of registry metadata records.

Although the metadata standards for registry developers are improving, the discussion turned to interoperability of registries with third party software applications through the use of Web Services protocols and standards. Discussion focused on the relative merits of Web service interfaces, such as REST, SOAP, ATOM, but a more cohesive approach to this was needed.

Dan Rehak recommended that the e-Framework¹⁰ adopted by e-learning authorities in the UK, Australia, New Zealand and Netherlands could provide the clarity required. Furthermore, it was agreed that more work was required to align the many data models and formats adopted by various groups operating in the registry space.

More generally, it was agreed that a prioritized list of 'use cases' should be developed to guide and refine discussion of the interoperability frameworks required for a global network of registries. This work would need to be done in future workshops and involve people with skills in data modeling and standards work.

The discussion then turned to the governance, policy and stakeholder engagement strategies required to make the registry network truly global. Not unexpectedly, this session raised as many more questions than it answered.

The first question was who to include in the network. Australia is currently establishing the Australian National Data Service (ANDS) and the UK has the Joint Information Systems Committee (JISC) to provide governance and a sustainable funding in this area; however, there are few equivalents elsewhere. How for example would the diverse interests of stakeholders in the USA be addressed?

Indeed, how would the for-profit (Google, Yahoo etc.) and not-for-profit (OCLC etc.) information service providers be included, or excluded as the case may be?

What, if any, international institution would take on the governance of a global network of registries? If none stepped forward, then would we need a new one?

Concerns were also raised about the framework for collaboration moving forward if no one person or group were given overall responsibility. There was also the perceived danger that its scope could grow, and that there would be no implementation or follow through. This led to a discussion for the need of development resource in order to move the initiative forward as it was clear that a number of people need to meet regularly—not just to clarify ideas but also to implement them.

Some good news came out of this discussion. Given the potential national benefits of a global registry initiative, Lucy Nowell and Andrew Treloar¹¹ (for the Australian Government through ANDS) indicated that funding for participation in International workshops might be considered by their organisations if timely decisions could be made about the scope of the workshops and participants.

The final session addressed the strategy for moving the global registry initiative forward. To reiterate—on the basis of the NSF and Australian Government representatives to consider funding further activities—the following was decided:

- An interim coordination group will be formed: Jeremy Frumkin to coordinate the US stakeholder involvement; Ann Apps to coordinate the UK stakeholders; Adrian Burton to coordinate Australian stakeholders.
- The coordination group will:
 - seek opportunities to publicize the initiative at international conferences and events (e.g. Open Repositories 2008)
 - follow up the idea of proposals to NSF and ANDS for workshop funding in 2008
 - distribute a stakeholder list for this group through the current wiki site hosted by Oregon University
 - write discussion papers, articles and reports detailing the initiative
 - establish working groups to address specific tasks and issues where relevant
 - write a detailed meeting report (this document).

Attendees

Jeremy Frumkin, Oregon State University

Adrian Burton, APSR

Todd Carpenter NISO

Kris Carpenter Negulescu, Internet Archive

Dan Rehak, eFramework

Dave Carpenter, NSF, NSDL

Lucy T. Nowell, National Science Foundation (NSF)
Ann Apps, Information Environment Service Registry (IESR)
Katherine Kott, DLF Aquifer
Andrew Treloar, ARROW/Australian National Data Service
Andreas Stenaus
Chris Blackall, APSR
Margaret Henty, APSR

Apologies

Herbert van de Sompel

¹ <http://ockham.org/>

² http://pilot.apsr.edu.au/wiki/index.php/ORCA_Registry

³ <http://iesr.ac.uk/>

⁴ <http://www.dlfaquifer.org/>

⁵ http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503141&org=OCI&from=home

⁶ Note that the meeting notes are not verbatim, but are summarized to give the reader a sense of the overall meeting and the issues it covered.

⁷ <http://www.imsglobal.org/>

⁸ <http://www.nla.gov.au/wgroups/ISO2146/>

⁹ <https://pilot.apsr.edu.au/cosi/about.php>

¹⁰ <http://www.e-framework.org/>

¹¹ <http://andrew.treloar.net/>