

Global Registries Initiative, Workshop Report 5-6 November 2009, Canberra

The Global Registries Initiative partners convened a workshop at the Australian National University (5-6 Nov. 2009) to discuss a common operational framework of the Initiative, to establish a model of shared governance, and to outline the scope and functions of services to be offered.

The workshop was divided into four sessions over two days. A workshop agenda and a discussion paper were circulated beforehand.

The agenda listed the specific issues and questions to be addressed in each of the four sessions. The structure of this workshop report follows the listed agenda items as discussed in each session. Beneath each session heading are the questions that were addressed, and beneath these are the notes of the discussion. Each session has a listing of the tasks that were identified.

Attendees

<i>Name</i>	<i>Affiliation</i>	<i>Email</i>
Jeremy Frumkin	Ockham Initiative	frumkinj@u.library.arizona.edu
Terry Reese	Ockham Initiative	terry.reese@oregonstate.edu
Vic Lyte	IESR	vic.lyte@manchester.ac.uk
Tim DiLauro	Johns Hopkins	timmo@jhu.edu
Adrian Burton	ANDS	adrian.burton@ands.org.au
Andrew Treloar	ANDS	andrew.treloar@ands.org.au
Chris Blackall	ANDS	chris.blackall@ands.org.au
Monica Berko	ANDS	monica.berko@ands.org.au
Heather Grant	ANDS	heather.grant@ands.org.au
Scott Yeadon	ANDS	scott.yeadon@ands.org.au
Nick Nicolas	Rapporteur (Link Affiliates)	opoudjis@optusnet.com.au

Day 1 Session 1

Vision

This session focused on where the Initiative has come from and where it is headed.

Suggested agenda items were:

- Revisit and refine the vision and goals of the GRI
- Discuss the business focus of the GRI including the granularity and coverage of the records provided by participating registries
- Discuss the functional scope of the GRI
- Discuss a sustainable business model for the GRI and what operational agreements might be needed to maintain it
- Discuss governance models, interim and long-term as also a possible organizational 'home' for the GRI

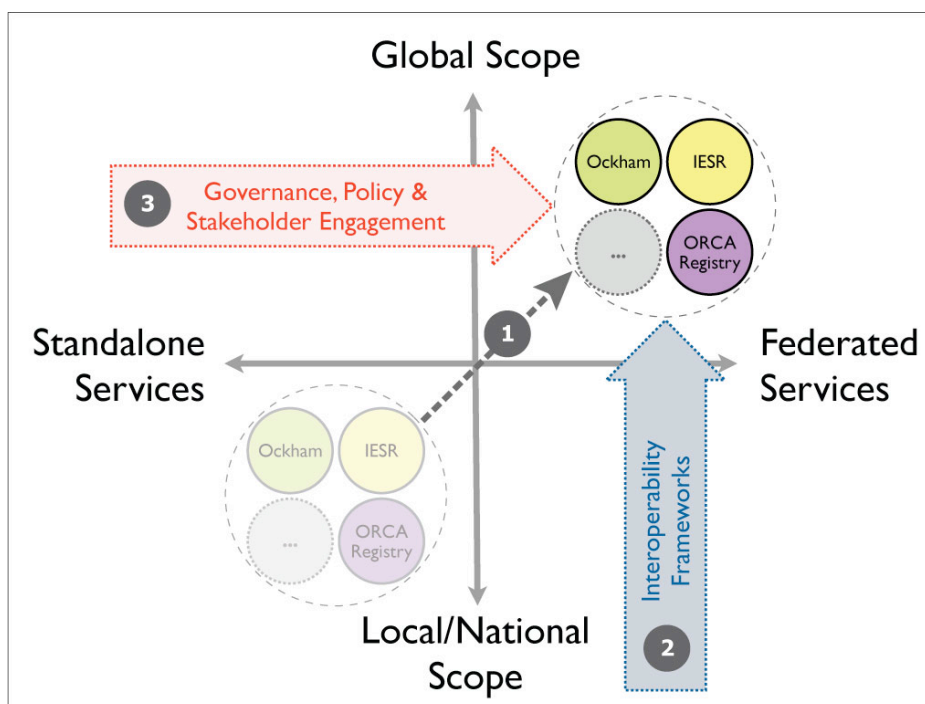
Notes

Adrian Burton opened the meeting by welcoming the delegates and providing a brief overview of how the GRI started and where it's headed (if all goes well). He used a slide (see figure 1) to illustrate the point that national registries (such as ANDS and IESR and Ockham)

need to address governance and policy issues and share a common interoperability framework if they are to be truly 'global' in scope and reach.

Adrian pointed out that registering research data is best performed at the local level, close to where the data is generated, but searching for scientific data is best done at the global level, giving the most effective discovery results. Or, as Adrian frames it: *register locally—discover globally*. The challenge therefore is to ensure that local, national and regional research collections can be made visible to the public at the global level. Google search already enables users to find this type of information; however, a great deal of information about data collections is still part of the 'dark web' and requires additional effort and infrastructure before it can become publically visible.

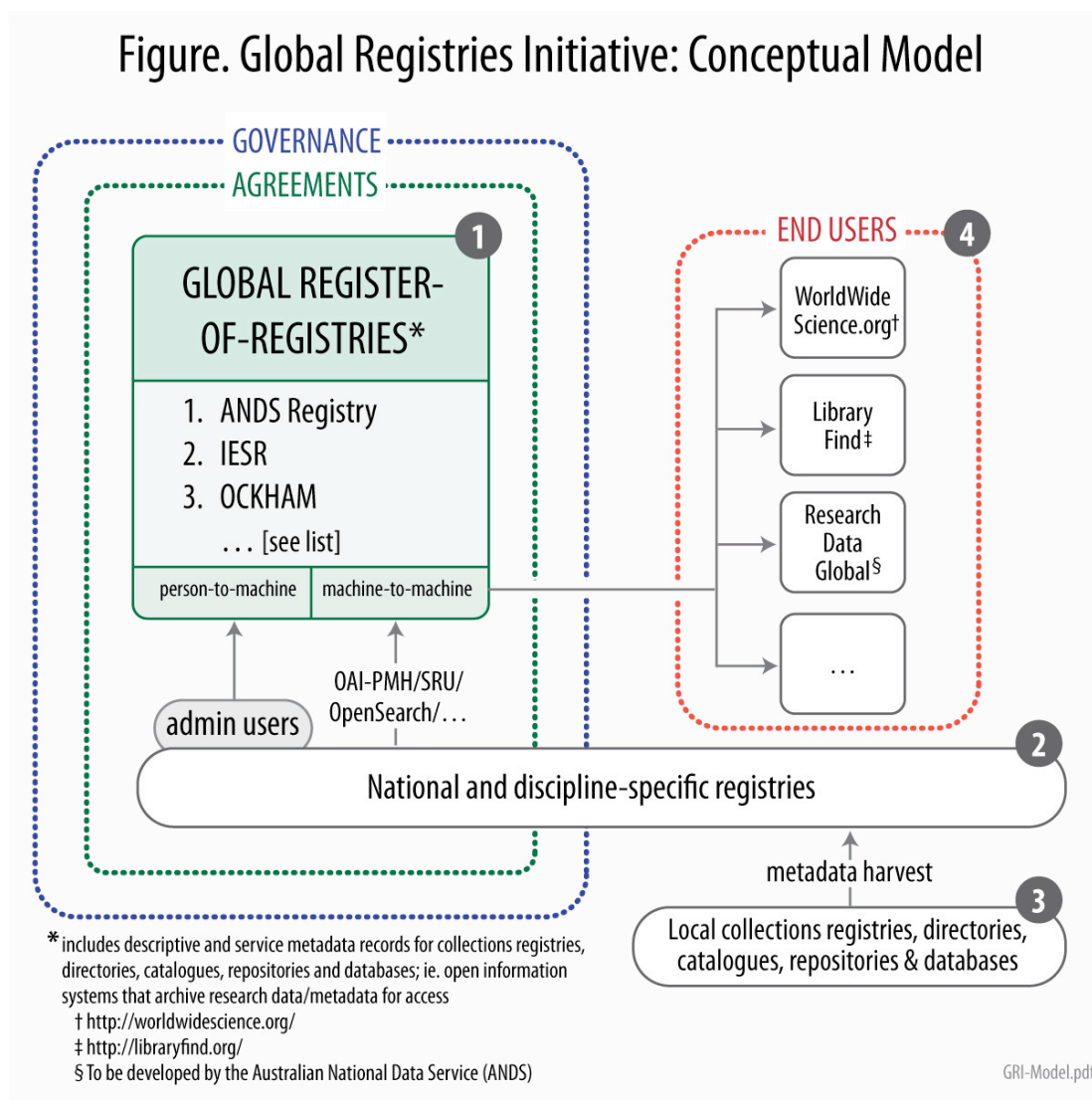
Figure 1.



The group agreed that it is not feasible to establish a single registration body to register the world's research data collections, but it is possible for this to be done at the international level by aggregating collections metadata from national registries, such as ANDS, IESR and OCKHAM are doing, and disciplinary registries such as arxiv.org, SSRN, and GEOSS.

The group reiterated the previous consensus on the architecture of the GRI in that metadata information about national and disciplinary registries, and their machine-actionable services, could then be made available at higher level of aggregation in a global 'register-of-registries'. Thus the articulated GRI 'vision' is of a federation of national and disciplinary registries, with an interoperable core of services based around a common 'register-of-registries' (see figure 2 below).

Figure 2.



The group discussed the functional elements of the conceptual model and their interrelationships (as defined in the discussion paper circulated before the meeting).

1. **Register-of-registries:** A registry instance that lists the registries that conform to the agreements established by the GRI; it is designed solely for the use of registry administrators and owners
2. **International, national and discipline-specific collection registries:** Registries, and other appropriate information systems, that gather collection and services metadata from research collections or downstream metadata aggregators; in this model these registries are responsible for gathering collection descriptions from their given 'jurisdiction', carrying out the agreed level of quality assurance for metadata, and exposing this information via GRI-compliant interfaces and information models.
3. **Local collections registries, directories, catalogues, repositories and databases:** Registries, and other appropriate information systems, that either hold research collections or who aggregate collection descriptions. Ideally they will supply collection descriptions to a national or discipline-specific registry (#2 in the conceptual model). In this model, this category of systems is typically an institutional, or discipline-owned, data facility, repository or portal.

4. **End-user data access points and discovery portals:** These are independent information systems that access machine-to-machine services information from the register-of-registries (#1 above) to provide new 'products'. In this model, we assume that an independent discovery service would be systems like [LibraryFind](#), [WorldWideScience.org](#), or [Research Data Global](#)¹.

The group paid particular attention on teasing out how the data access points and discovery portals (#4 above) would use the GRI. This was important because these are the systems/applications that will ultimately benefit from the GRI enabling infrastructure.

In addition to confirming the basic architecture and functions of the GRI, the group also made a key decision to extend the GRI 'vision' to enable the registration of registries (and like systems) that hold information about learning resources. This was to take into account the broad requirements of JISC and the role of the National Digital Science Library of which Ockham Initiative is linked.

1. Review the vision statement of GRI (as articulated on the website) in light of discussions during the workshop regarding business scope and objectives [Adrian Burton & Monica Berko]

Day 1 Session 2

Business Focus

This session took the outcomes from discussions in the previous session and focused on the decisions required to progress the 'register-of-registries' (#1 above) and the desired downstream services. The suggested agenda items were:

- Interim infrastructure for register-of-registries and [globalregistries.org](#) web site
- Interim infrastructure for GRI discovery portals (#4 above)
- Consider how collections granularity and coverage can be recorded in metadata so that the merging of registries provides a whole that is greater than the sum of its parts.
- What exchange schema(s) should be used by participating registries so that all metadata relevant to the scope of the GRI is interchangeable including relationship metadata (within and between collections, services and parties/agents)
- What metasearch and harvest protocols must be supported by each participating registry and to what level (e.g. support for incremental harvesting)
- What are the minimum requirements for a participating registry in the areas of (a) purpose (b) responsible parties (c) metadata quality (d) operational/service standards

Notes

Based on the discussion in the previous session, it was agreed that it was important to better contextualise the use of registries, such as GRI.

The basic information model that GRI supports is one that binds service descriptions (ie. Web Service) to collections metadata, thus linking the end-user discovery of collections to their (re)use via services. The key beneficiaries of the GRI are not only end users seeking information about collections, but also down stream metadata aggregators and intermediaries (portals, applications, catalogues, mashups and so on).

¹ <http://services.ands.org.au/global/orca/rda/>

While developing discovery portals is out of scope for GRI itself, it assumes that the GRI partners are free to build end-user services enabled by the GRI registry of registries. Indeed, LibraryFind already serves as a demonstrator of how the GRI can be used, as does the Research Data Global demonstrator developed by ANDS.

Tim DiLauro also pointed out that the Data Conservancy initiative offers potential use cases for GRI; for example, it could re-expose content via GRI for consumption. [\[Tim, could you provide more detail here?\]](#)

Although it was confirmed that the coverage of the GRI should include scholarly information and learning resources as well as research data collections, it was agreed that there was no point in the GRI duplicating existing global library systems such as the OCLC Registry. The OCLC Registry has been as successful as it is because it controls its own client 'ecosystem'.

The GRI similarly needs to establish which 'real estate' it occupies in the research data landscape, establishing its context and value to end-users. This is a challenge, as the repository and data management space is continually evolving, especially given the recent funding inputs from NSF, JISC, and the Australian Government via ANDS.

2. LibraryFind service to be prepared for demonstrations as soon as possible, and to be robust for deployment [Jeremy Frumkin].
3. ANDS to develop two end-user applications (see #4 above, Research Data Global and a 'see-also service' in the Research Data Australia) using the register-of-registries [ANDS software development team]
4. NSF Data Conservancy to identify potential data providers and consumers of the register-of-registries within its stakeholder community [Tim DiLauro].

Functional Scope and Application

It was agreed that the GRI, as a machine-actionable global register-of-registries, will promote targeted harvesting of registries as a primary use case. This necessitates a shared descriptive model of registries, with both human- and machine-actionable descriptions of collections to enable collection discovery. Until now, data interchange modelling was done on an ad hoc basis; however, it was agreed that this was unsustainable for GRI.

Based on this agreement, the group considered that it establish more concrete use cases for how the GRI Registry may be used, and what type of information registered with GRI would be useful to end users, service providers, and mashup communities. These use cases will guide the refinements of the agreements between registries, as to what kind of information they will expose.

The group then dealt with the issues around the quality assurance (QA) of the data exchanged within the GRI partnership. While it was agreed that a good level of QA would be ensured for the register-of-registries, it was acknowledged that the access and QA rules associated with 'collections' cannot be policed by GRI, and remain the domain of the collections themselves, realised through their services layer. There was extensive but inconclusive discussion on this topic. It was decided to hold discussion to the second session on day two.

It was also decided that the GRI, as transparent infrastructure, would not provide a 'discovery service' itself, but would endorse compliant service providers to do so, linking to them on the GRI web site in the first instance. Different options to do this were discussed but a clear solution was not identified. Nevertheless, the group agreed that the services so

linked would need to have proven utility for GRI users, as established by the commissioned use cases.

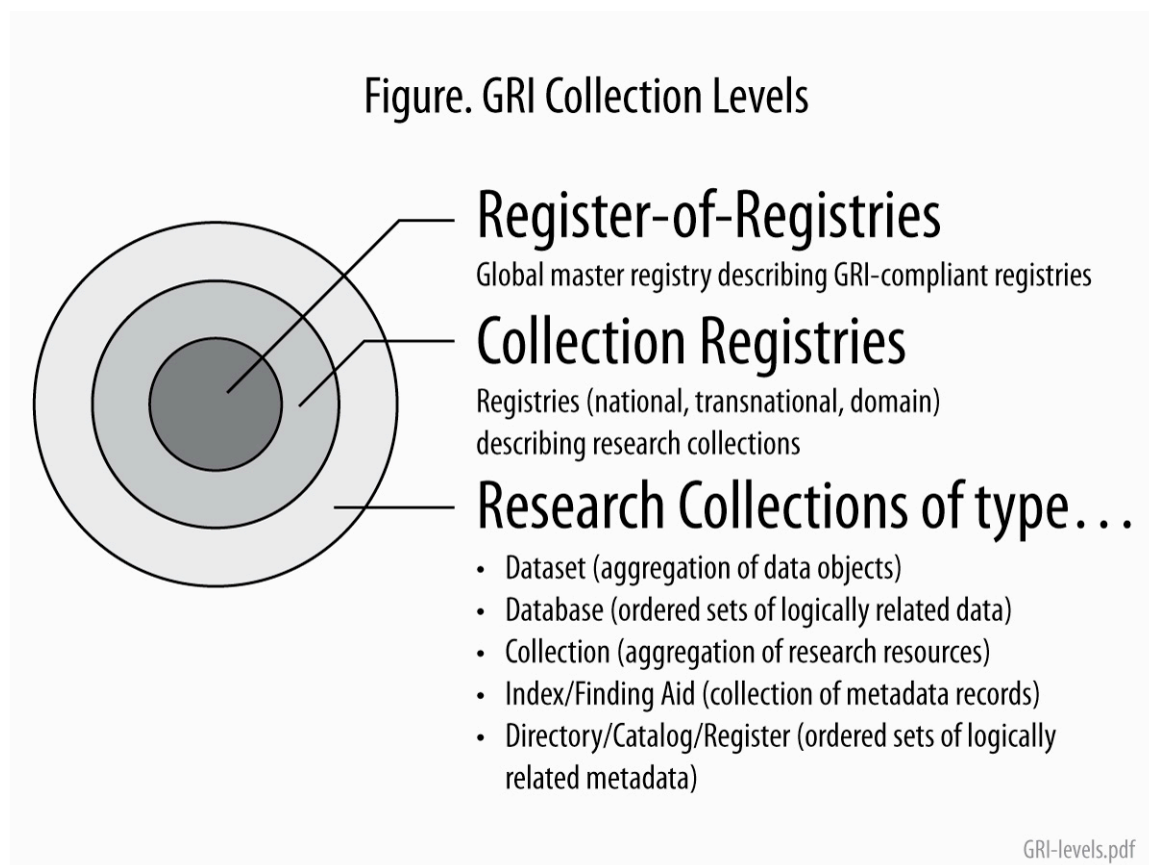
It was further agreed that the interim GRI Management Group would be investigate and review the 'discovery' options for the GRI. Similarly, it was agreed that the interim GRI Management Group would establish an Architecture Working Group to undertake some of the modelling work required to substantively address the GRI functional architecture.

5. ANDS to contribute use cases towards GRI [Chris Blackall & Monica Berko].
6. NSF Data Conservancy to contribute use cases [Tim DeLauro].
7. IESR to provide two use cases from previous survey of community requirements [Vic Lyte].
8. Develop an information/data model for the register-of-registries (required for action 21) [Chris Blackall & Monica Berko].
9. Develop UML sequence diagram to clarify functional interactions between systems [To be generated by architecture working group i.e. Chris Blackall & Tim DiLauro].
10. The interim GRI Management Group will review discovery requirements over the next 12-18 months [GRI Management Group].

Registry information model

As established previously, the GRI will expose its federated hierarchy of registries in machine-actionable form; nevertheless, the entities in this hierarchy and their granular relationships have yet to be sufficiently defined. A graphic (figure 3) suggesting a nominal hierarchy was included in the discussion document.

Figure 3



The need for definitions of entity within the GRI functional hierarchy was accepted and the group discussed some of the options (based on figure 3).

It was agreed that without a consistent and logically valid model of GRI entities, there is a risk of confusion and category errors. Hence, it was agreed that an information model for GRI entities would help to address this problem (see task 11).

It was also agreed that this task may require authority control mechanisms, such as controlled vocabularies, to harmonise the categorisation of collection registries. At this point, there was a lengthy discussion about the need (or not) to define and use registry types (e.g. repository, directory, index, finding aid etc. as defined in ISO2146/RIF-CS and the Dublin Core Collection Application Profile) and require their use by GRI partners. The consensus was not to impose this requirement on GRI partners; however, these entities would still need to be defined in order to better communicate the function and scope of the GRI itself.

That said, the definition of what counts as a 'collection' for the GRI will need to follow whatever the community wants a collection to be and a considerable degree of heterogeneity will need to be tolerated.

11. Develop definitions of system-of-systems and register-of-registries, and diagram levels of organisation and interaction between functional entities [Chris Blackall & Tim DiLauro].

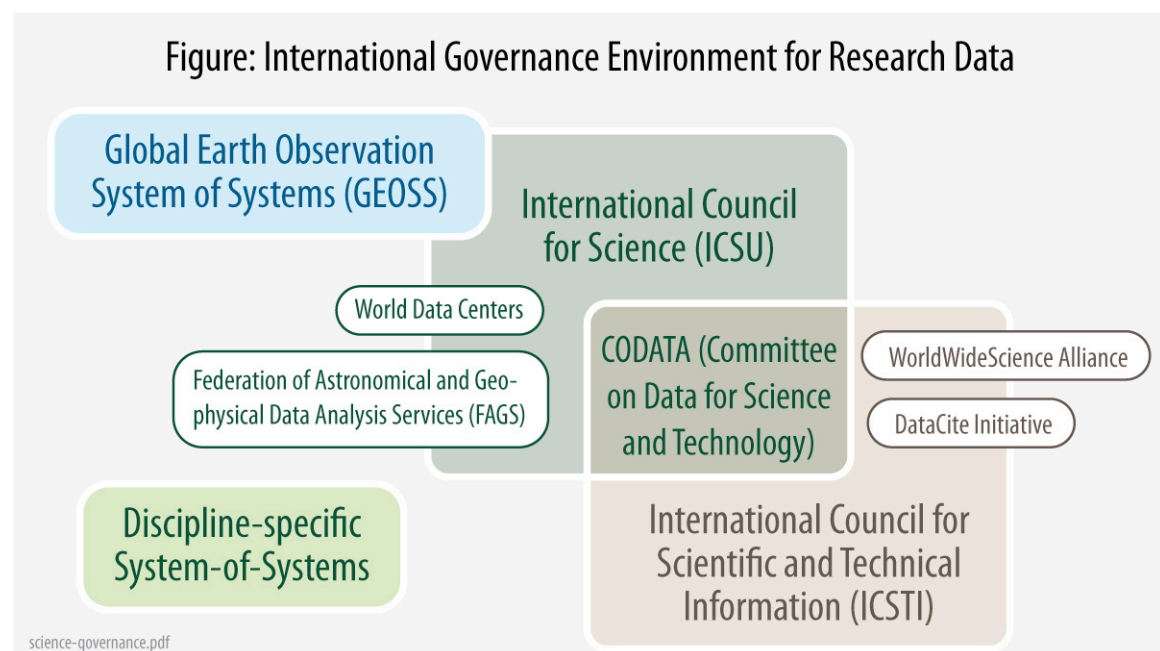
Sustainable Business Model & Governance

It was agreed that GRI requires international governance with stakeholder engagement. In the short term, it needs interim agreements between the GRI partners in order to get

established. As previously stated, it was decided that an Interim Management Group will be formed.

In terms of establishing the right governance model for ANDS, Chris Blackall spoke to a diagram (figure 4) and the relevant section of the meeting discussion document on governance. He mapped out the key science data governance bodies and their interrelationships, pointing out that the overall governance environment is becoming increasingly fragmented and lacking clear policy goals and sustainable funding.

Figure 4.



Nevertheless, for GRI to achieve economic and functional sustainability, it was agreed that once GRI has established its own interim governance arrangements, it needs to align with global science governance bodies (ICSU, CODATA and ICSTI), standards bodies, and funding bodies. It was argued that the GRI would gain the following advantages from such alignment:

- legitimacy within the science community
- potential source of operational funding
- strategic tie-ins with standards bodies and initiatives
- opportunities to establish GRI 'brand' recognition
- alignment with complimentary initiatives such as GEOSS, WorldWide Science Alliance, DataCite and so on

Following this discussion, the following tasks were identified:

12. Establish interim GRI management group to commission and approve future work on architectures, governance etc. [Interim GRI Management Group]
13. Develop a strategy to contact potential candidates for registration in the register-of-registries [Interim GRI Management Group]
14. Write a position paper addressing issues related to association with external scientific governance bodies, sustainable funding, standards conformance and 'market' positioning of the GRI within global cyberinfrastructure [Chris Blackall & Monica Berko]
15. Ongoing monitoring of relevant NISO standards activity [Jeremy Frumkin].

The group then discussed the desired timeframe for GRI to extend its current (three partner) membership. GRI is pursuing collaboration with COAR (Confederation of Open Access Repositories), but COAR are still waiting for their funding to come online. JIRO (Japanese Institutional Repositories Online) is another potential collaborator. No current contact point has been identified within ICTSU/CODATA; and it was agreed that this needs to be revisited. The stated aspiration was to sign up three more partner bodies by the end of 2010, but no one was allocated this task.

Furthermore, because of the ambitious scope of the GRI, it was recognised that any rollout of its 'services and collections' model needs to be staged to remain pragmatic. Hence, GRI will look at 'low hanging fruit' to include in its initial drive for data providers. Examples include the Biodiversity Collections Index and the IVOA, which are registries-of-registries in their own rights. These groups already have service registries, are leaders in their domains, and have demonstrated progress in line with the goals GRI is pursuing.

16. Contact to be made with Biodiversity Collections Index via Atlas of Living Australia [Adrian Burton].
17. Contact to be made with IVOA, Encyclopaedia of Life, Biodiversity Heritage Library [Tim DiLauro].

Day 2 Session 1

This session focused on the data and information mapping between the registries and the harvesting issues encountered during the recent harvest of data from the three registries into the pilot global collections registry.

Suggested agenda items:

- Overview of the IESR and RIF-CS schemas and sample records taken from each registry
- Review of the mapping used to transform IESR records to RIF-CS format and issues that need to be resolved
- Required process and systems for updating the register
- Harvesting and metasearch technicalities

Services

The first issue to be addressed was the fact that the GRI partners have been using (slightly) different protocols, schemas and payloads.

It was agreed that the register-of-registries will support full OAI-PMH (with deleted records and incremental harvest), and the OpenSearch protocol.

Each 'conformant' downstream registry (ie. #4 in figure 2) should support OAI-PMH, and optionally OpenSearch also. These endpoints will be recorded as distinct service records in the metadata registered in national and discipline registries (ie. #3 in figure 2).

It was agreed that participating registries that no longer conform to GRI membership rules should be removed from access through GRI.

18. Formulate minimum metadata requirements for participating registries, including OAI-PMH implementation [Monica Berko]

Metadata Exchange format

The meeting then went onto discuss the technical options for exchanging metadata records between GRI partners.

There were two options discussed here: 1) to use a modified version of the IESR Schema; or 2) the RIF-CS instantiation of the ISO2146.

After due consideration, it was agreed that the GRI would use a modified version of the IESR Schema (taking into account the fact that IESR is currently shifting to a simplified version). The reason for this decision was that the IESR-based application profile would capitalise on metadata that is already being exposed by partners.

Consequently, the GRI will develop a new metadata application profile based on the IESR schema. The data model will start with the current version of IESR's schema, and will not introduce added dependencies on other schemas, or on internal IESR updates to their own schema. Most of the work here is likely to be on schema profiling and optimisation, rather than on modifying the underlying IESR information model.

Some revisions may be entertained; however, the group thought that it was important to release the candidate application profile as soon as possible, preferably by March 2010.

19. Analyse and review the IESR information model for applicability to GRI [Scott Yeadon, Terry Reese and nominee from IESR].
20. Develop metadata application profile and (XML) data interchange format [Scott Yeadon, Terry Reese and nominee from IESR]

Day 2 Session 2

The final session focused on developing a written strategy/plan that will enable the group to turn the unified vision into a reality. The topics to be covered were:

- Interim governance documentation, policies, meeting schedule
- Communications Plan
- Technical Agreements for participating members and the process for updating them
- Maintenance of the Web Site and GRI Register

Documents

Much of this session was taken up with rehashing some of the topics that had not been sufficiently addressed during the previous day of the Workshop; hence, the goals of drafting some of the key governance, policy and operations documents was not achieved, nor was a timetable set for achieving this outcome.

Resources

Despite the absence of drafting the desired documents, the group made a number of key decisions needed to make progress. These are listed below:

21. Provide ongoing secretariat resources and support [Monica Berko assigned to this task].
22. Establish software instance of the register-of-registries, on behalf of GRI (as per action 8) [ANDS software development team].
23. Provide infrastructure for GRI wiki [Tim DiLauro].
24. Provide hosting for current GRI web site [Adrian Burton & Monica Berko]
25. Circulate meeting notes (this document) [Chris Blackall, Monica Berko]

Conclusion

The consensus of the meeting was that the GRI partners had made solid progress towards what is unquestionably a very ambitious initiative. The group acknowledged that it is somewhat ahead of the wave in this area, but it was heartened by the fact that other leading international science and data curation groups have independently established their own 'system-of-systems' (comprised of federated systems of registries and repositories).

The next challenge will be to establish and document the governance, policy and operational arrangements for the GRI. Although these topics were discussed extensively, it was acknowledged that the interim GRI management group would need to make key decisions in this area fairly soon.

A date and time for a follow up meeting have yet to be set.