

Aligning Repository Networks and the Confederation of Open Access Repositories (COAR)

Kathleen Shearer

Abstract

Research is becoming increasingly international. Many of today's greatest challenges such as climate change, poverty, and health are global in nature and must be addressed in collaborative ways by researchers across regional and disciplinary boundaries. In this environment, research infrastructure should be connected, networked and developed to reflect the evolving needs of the research community.

COAR (Confederation of Open Access Repositories) is an international organization with members from over 35 countries on 5 continents. COAR's mission is to raise the visibility of research outputs through a global network of repositories. We are active on two levels: (1) At the practical level, we support communities of practice around areas of importance for our members mainly in terms of best practices, interoperability and monitoring trends in the repository landscape and (2) At the strategic level, we aim to facilitate greater alignment of regional and national repository networks around the globe.

In March 2014, the Confederation of Open Access Repositories (COAR) launched a major initiative to align repository networks across the world. The aim of the project is to establish a mechanism for ongoing dialogue between repository networks. Strategically, it will give the repository community a stronger global voice and raise the visibility of the role of repositories as critical research infrastructure. It will act as leverage for local initiatives and demonstrate that networks are aware of and operate consistently with international trends. At the practical level, this activity will allow repository networks to discuss and adopt best practices for metadata standards, vocabularies and services.

Keywords: COAR, Institutional Repositories, Open Access Repositories

1. Introduction

Open access repositories are becoming key components of the research infrastructure. They represent an important content layer within the research infrastructure system and fulfill a number of different roles. To start with, repositories provide access to the products of research to researchers and the world. They also reflect an emerging commitment by research institutions towards the stewardship of

the research outputs they produce. Furthermore, repositories are also becoming an important source of information for governments, funding agencies and institutions about the impact of the research they support.

Many regions around the world are investing in the development of repository networks. These networks have evolved in their specific local contexts and currently differ in a number of ways. However, the real value of repositories is when they are interconnected to provide unified access to research materials for researchers around the world and allow



10th International CALIBER-2015
HP University and IAS, Shimla, Himachal Pradesh, India
March 12-14, 2015
© INFLIBNET Centre, Gandhinagar, Gujarat, India

us to aggregate, data mine, create new tools and services, and generate new knowledge from this content. . To achieve this, there must be some level of alignment across repository networks.

Repository networks can be aligned on a number of levels:

- ▶▶ Policies (and laws): harmonizing requirements across policy environments
- ▶▶ Technical standards: adhering to common metadata, standards, and vocabularies
- ▶▶ Services: developing and supporting shared services
- ▶▶ Strategic alignment: Dialogue, adopting common vision and strategy, working with other stakeholders (funding agencies, CRIS systems, publishers, software platforms) to ensure interoperability with policies and systems

2. Benefits and Challenges of Aligning Repository Networks

As research becomes increasingly global, distributed and cross disciplinary, repository infrastructures must mirror the needs of the research community and enable researchers, regardless of location or disciplinary practice, to access research outputs worldwide. All efforts should be made to avoid silos, which act as barriers to the use of content.

Aligning repository networks will help us collectively to advance towards the vision of a seamless global knowledge infrastructure. It will enable the exchange of data between repositories and support the creation of new services such as disciplinary portals or text mining. Through the adoption of common standards for usage (metrics) and common vocabulary elements, aligning repository networks will allow governments and funding agencies to gather more

uniform information about the impact of the research they fund. In addition it will enable networks to learn from each other allowing the global community to progress more quickly leading to cost synergies by preventing duplication of work across networks.

There are a number of challenges associated with aligning repository networks. There is significant diversity across regions in terms of implementation speed and availability of resources. Networks don't share a common directive and have been deployed to support differing mandates and requirements. In addition, differences in language and wide geographic distribution present challenges to working together and identifying common approaches.

Nevertheless it was agreed that there are some important areas where repository networks currently intersect and should be aligned, and there will continue to be issues of common interest as networks evolve.

3. Current State of Repository Networks

There are numerous national and thematic repository networks around the world, which link repositories with each other. These have evolved based on unique requirements and mandates; are at different stages of development; and reflect varying levels of integration. Some national networks, such as in the UK, Argentina, and Spain, are very cohesive and have a number of robust services supporting their repositories. Others are less developed and unified, and revolve more around a community of practice for repository managers. Broadly speaking, repository networks can be characterized as having one or more of the following aspects: community of practice, adoption of common standards for metadata and vocabularies, centralized harvester, catch-all re-

positories for orphan publications, and other value added services.

In addition to national and thematic networks, regional repository networks are being developed to connect repositories across national boundaries. In Latin America, LARreferencia is a network of repositories in nine countries: Argentina, Brazil, Chile, Colombia, Ecuador, and México, Peru, Venezuela and El Salvador. The initiative began as a project funded by the Inter-American Development Bank (IDB) and is now managed by CLARA, the organization that manages the high-speed network in Latin America. LARreferencia develops strategies, maintains a centralized harvester and promotes common standards across Latin America.

OpenAIRE is a project funded by the European Commission (EC) to develop repository infrastructures in the European Union. OpenAIRE aggregates the research output of EC-funded projects and makes them available through a centralized portal. OpenAIRE repositories adopt common guidelines so that content can be aggregated into the central

portal. OpenAIRE is also developing a number of value added services that will enable users to use content in new and innovative ways including text mining, and statistics and reporting tools, which enable users to better track funded research outputs.

Another recently launched initiative, that will likely have an impact on the global repository environment is the SHARE project in the US. SHARE (SHared Access Research Ecosystem) is a joint effort supported by the Association of Research Libraries (ARL), the Association of American Universities (AAU), and the Association of Public and Land-grant Universities (APLU) to strengthen efforts in the US to identify, discover, and track research outputs. SHARE, still in the early stages of development, aims to develop a network with three layers: a distributed registry layer for publications and research data; a discovery layer to improve access to content; and an aggregation layer to facilitate data and text mining and other value added services. The first step, the registry layer, will require the adoption of standard metadata to identify and process research release events.

Table 1: Policies

Policy Elements	Current State
Method for open access: discipline-based repositories, institutional repositories, OA journals, or all.	<ul style="list-style-type: none"> • Most policies allow researchers to use all avenues to provide open access • NIH and other health funding agencies require deposit into PubMed Central • Peru and Argentina require deposit into institutional repositories • Others require OA by repository or journal, but metadata must be deposited into
Embargo periods	<ul style="list-style-type: none"> • Most common embargo period for access is 12 months • There are a few policies that allow longer embargoes (usually humanities or social sciences) • RCUK requires six months for articles in STEM • Often require immediate deposit, with embargo access period

Content types	<ul style="list-style-type: none"> • Journal articles (and conference proceedings) • There is a growing number of mandates covering research data, but requirements often differ • Monographs- only a few
Versions: author's final manuscript	<ul style="list-style-type: none"> • Author's final manuscript is most common • Final published version
Re-use requirements	<ul style="list-style-type: none"> • Most policies simply require that the content is available free of charge • RCUK: Search for and re-use the content of published papers both manually and using automated tools (such as those for text and data mining)

Table 2: Technical Interoperability

Technical Issue	Current State
Metadata standards	<ul style="list-style-type: none"> • Most repositories are at minimum OAI-PMH compliant and the protocol has become the baseline for repository interoperability. All major repository software platforms nowadays feature a default OAI-PMH protocol, making it the easiest interoperability protocol to implement and is the starting point for repository interoperability. • In Europe, many repositories adhere to DRIVER or OpenAIRE guidelines. A number of repositories in Latin America have also implemented DRIVER and OpenAIRE guidelines. • Repositories in UK are requested to adhere to RIOX Guidelines. • In US, SHARE is developing metadata standards for the Notification System (currently assessing RIOXX) • NISO proposed Open Access Metadata and Indicators
Usage statistics	<ul style="list-style-type: none"> • No official adoption of usage statistics by repository networks, although there are a number of projects and guidelines being developed listed below • Repositories rely on default usage statistics tools built in their repository platform • Knowledge Exchange Usage Statistics Guidelines • OA-Statistik • PIRUS/IRUS-UK, and SURE initiatives all provide methods or mechanisms to collect cross-repository usage statistics.
Standard vocabularies	<ul style="list-style-type: none"> • The controlled vocabulary "info:eu-repo", developed by DRIVER and OpenAIRE European projects has been widely adopted in Europe.
Author identifiers Link individual authors to their publications	<p>Several initiatives exist:</p> <ul style="list-style-type: none"> • ORCID seems to have the most traction in North America • AuthorClaim, a scholar may "claim" his/her associated publications, which can then be jointly displayed. • ResearcherID (Thomson Reuters) • Author Identifier (Elsevier)

<p>Persistent identifiers</p> <p>Persistent identifiers for digital objects.</p>	<p>Several initiatives exist:</p> <ul style="list-style-type: none"> • The Digital Object Identifier (DOIs) • Handle system- a technology specification for assigning, managing, and resolving persistent identifiers • ARK (Archival Resource Key) Identifiers • PURLs • DataCite is more precise in that it is specifically designed to assign persistent identifiers for datasets
<p>Cross-system transfer</p> <p>Enabling multiple-deposit or transferring content from one system to another.</p>	<ul style="list-style-type: none"> • The SWORD protocol and the Open Access Repository Junction (UK) were both designed to support cross-system content transfer. • For compound digital objects, the Object Reuse and Exchange (OAI-ORE) specification focuses on creating systemized ways to move bundles of objects from one system to another.

Table 3: Services

Service	Description
<p>Publisher policies</p> <p>Discovery/ harvesting</p> <p>OAI-PMH is the basis for this approach, with a minimal number of additional fields layered on top of OAI-PMH.</p>	<p>Sherpa-Romeo: Publisher copyright & self-archiving policies</p> <p>CORE: CORE (CONnecting REpositories) aims to facilitate free access to scholarly publications distributed across many systems. As of today, CORE gives you access to millions of scholarly articles aggregated from many Open Access repositories. Managed by Open Knowledge Institute, UK</p>
	<p>DRIVER: Access the network of freely accessible digital repositories with content across academic disciplines with over 3,500,000 scientific publications, found in journal articles, dissertations, books, lectures, reports, etc., harvested regularly from more than 295 repositories, from 38 countries.</p> <p>LaReferencia Portal: Harvests content from nine Latin American countries</p> <p>EBSCO Discovery Service: Institutional Repositories (IRs) can be directly loaded into EDS so that they can be fully searched alongside all other EDS resources/ content.</p> <p>National portals/harvesters</p>
<p>Repository directories</p> <p>Improves visibility of repositories</p>	<p>OpenDOAR: an authoritative directory of academic open access repositories.</p> <p>ROAR: Registry of Open Access Repositories</p>
<p>Cross-system transfer</p> <p>Enabling multiple-deposit or transferring content from one system to another.</p>	<p>The SWORD protocol and the Open Access Repository Junction (UK) were both designed to support cross-system content transfer.</p> <p>For compound digital objects, the Object Reuse and Exchange (OAI-ORE) specification focuses on creating systemized ways to move bundles of objects from one system to another.</p>

Monitoring research output	OpenAIRE: harvests and connect publications to the related EC FP7 grant agreements SHARE Notification System: In this initial stage, SHARE will focus on the development and deployment of a common notification system that notifies any interested stakeholder of the release of research results.
Other services in development	Text mining: "The global research community generates over 1.5 million new scholarly articles per annum. 1 As the recent Hargreaves report into 'Digital Opportunity: A Review of Intellectual Property and Growth' 2 highlighted, text mining and analytics of this scholarly literature and other digitised text affords a real opportunity to support innovation and the development of new knowledge." (JISC 2014) Linked data: Best practices for publishing and connecting structured data on the web. The goal is to create the "semantic web". (Berners-Lee 2009)

4. COAR's Aligning Repository Networks Initiative

In March 2014, the Confederation of Open Access Repositories launched a major initiative to align repository networks. COAR is an international organization with members from over 35 countries on 5 continents.

The aim of the initiative is to establish a mechanism for ongoing dialogue between repository networks. Strategically, it will give the repository community a stronger global voice and raise the visibility of the role of repositories as critical research infrastructure. It will act as leverage for local initiatives and demonstrate that networks are aware of and operate consistently with international trends. At the practical level, this activity will allow repository networks to discuss and adopt best practices for metadata standards, vocabularies and services.

In many cases there are already bilateral discussions taking place. The aim is to build on these existing activities and further expand them to involve other networks. In addition, in each priority area, there will be other key stakeholders that are critical for moving forward. Where appropriate, COAR will work closely with relevant stakeholders including

other national and thematic networks, allied organizations such as SPARC, RDA, COAPI, IFLA, as well as organizations maintaining infrastructure, services or developing repository platforms. In the coming weeks, COAR will develop a more concrete plan for moving the priority areas forward.

5. COAR-CASRAI Working Group on Interoperability of Open Access Repositories

As a first step towards Aligning Repository Networks, COAR launched an international, multi-stakeholder group to develop a strategy to ensure greater technical interoperability across repository networks and other platforms. The Confederation of Open Access Repositories (COAR) will be the convener of the working group, and CASRAI will facilitate the process of developing the strategy. Members of the working group will be representatives from major regional repository networks, EuroCRIS and CASRAI and will be committed to adopting the outcomes.

Greater interoperability across these metadata standards and guidelines will improve discovery and enable the development of more sophisticated cross-repository services such as tracking and monitoring research outputs, evaluation usage and impact or

content analysis through text mining. In addition, a shared understanding and common approaches to vocabularies and metadata will provide uniform information to governments and funding agencies about the products of funded research, and build confidence with users and stakeholders.

The aim of this working group is to develop a blueprint that will outline the steps needed to ensure greater interoperability across repository networks, and ideally with other related systems and actors as well (publishers, CRIS, etc.). The group began its work in Oct 2014 via a mix of teleconference and online collaboration tools and aims to have a completed plan by April 2015.

Members of the working group represent major repository networks, research administrative systems, and other stakeholder communities. (COAR, CASRAI, EuroCRIS, Jisc/UK, La Referencia, OpenAIRE, NISO Open Access Metadata and Indicators Working Group, and SHARE).

To date the group has identified two major use cases: (1) A funding agency is tracking research outputs of its funded research across several repository networks, and (2) An institution is monitoring the research outputs that are also scattered across a number of different repositories. The group is now in the process of reviewing each metadata element that contributes to fulfilling these use cases, with the aim of adopting common vocabularies and metadata schemas wherever possible.

5. Conclusion

Openness is about more than gaining access to knowledge. It is also "about the right to participate in the knowledge production process, driven by issues that are of local relevance, rather than research agendas set elsewhere or from the top down" (Chan

2014). And while open access has made tremendous gains over the last several years to improve access to research publications, APCs have become prevalent, with the consequence of further marginalizing researchers at institutions in the developing world who cannot pay the increasingly large fees attached to them.

However, COAR envisions a system to which all researchers can access and contribute to, regardless of geographic location or discipline; and where the knowledge created is assessed on its real value, rather than on the region from which it emerges or the so called "impact" of the journal in which it is being published.

To achieve this, we aim to collectively build and maintain a global system of repositories, that presents the important knowledge that created through research around the world. The only way we can achieve this is by working together towards a common vision.

References

Chan, Leslie. "What role for open and collaborative science in development?" *University World News*, September 5, 2014, Issue No:333. Available at: <http://www.universityworldnews.com/article.php?story=20140905132026576>

About Author

Ms. Kathleen Shearer, Executive Director of COAR, 1079 Atwater Avenue, Montreal, Quebec, Canada, H3H 1X5.

Email: m.kathleen.shearer@gmail.com