Promoting greater visibility and application of research through global networks of Open Access repositories

Towards a Seamless Global Research Infrastructure

Report of the Aligning Repository Networks Meeting, March 2014

Prepared by Kathleen Shearer
Executive Director of COAR

April 25, 2014

http://coar-repositories.org

Table of Contents

I. Introduction

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 developed to reflect the evolving needs of the research community.

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 us to aggregate, data mine, create new tools and services, and generate new knowledge from this content.[1] To achieve this, there must be some level of alignment across repository networks.

In Rome on March 20 & 21, 2014, the Confederation of Open Access Repositories (COAR) convened a meeting to develop a strategy to better align regional repository networks. At this meeting, delegates from Australia, Canada, China, Europe, Latin America and the United States identified several key elements that will be addressed immediately and agreed to work together on an ongoing basis on other issues to further align their repositories. This report provides a summary of meeting discussions and outlines next steps identified by participants.

A list of meeting delegates and their affiliations is available in Appendix 1.

[1] Please read COAR's "Case for Interoperability for Open Access Repositories" for a more detailed description of the value of interoperability: https://www.coar-repositories.org/files/A-Case-for-Interoperability-Final-Version.pdf

II. 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 repositories 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, LAReferencia is a network of repositories in nine countries: Argentina, Brazil, Chile, Colombia, Ecuador, and Mexico, 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. LAReferencia 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-compliant 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.

III. Challenges and Benefits 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.

Delegates agreed that aligning repository networks will help us to collectively 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.

Participants concurred that it would be impossible, and also undesirable, to seek to align networks across all areas and elements. Networks have been developed based on unique requirements and have evolved to fulfill the needs of their own individual regional context.

It was also acknowledged that there are some 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.

IV. Vision and Principles

Meeting delegates agreed on a common vision and principles that will guide the collaboration.

Our Vision

Researchers, regardless of location or discipline, have open access to the valuable content created through publicly funded research. In support of this, repository networks work together to provide seamless access to research outputs and adopt common practices that maximize the ethical re-use of content and development of value added services.

Our Principles

Openness: Open access to research outputs resulting from publicly funded research should be timely and user friendly.

Sustainability: Research outputs of long-term value should be accessible for current and future generations. In order to achieve this, research infrastructure should ideally be managed by long-term actors that have a mandate for stewardship and preservation.

Interoperability: Technological and semantic interoperability are key considerations for enabling access and re-use of content. Repository networks should adopt common standards across networks in support of interoperability.

Diversity: Diversity in approaches is beneficial. Diversity drives innovation and enables regions to respond to unique objectives and requirements.

Current Priorities for Alignment

Meeting delegates discussed three levels of potential alignment for repository networks:

- Policies (and laws): harmonize requirements across policy environments
- Technical and semantic standards: adopt common metadata standards, elements, and vocabularies
- Services: support, adopt, and develop shared services

As a first step, the following priorities were identified:

Policies	Limiting embargo periods: Any delay in the free availability of research outputs will curtail scientific progress and innovation. Repository networks will continue to emphasize that immediate access should be considered best practice. If embargos are to be imposed they should be a maximum of 6-12 months (depending on discipline).
Technological and Semantic Interoperability	Incorporate ORCID IDs into repositories globally Promote the use of persistent identifiers Standardize certain vocabulary elements that express important concepts including, but not limited to: embargo periods, access and re-use conditions, funder IDs Establish and adopt of common approaches to collecting usage data and impact measures Implement protocols that enable data exchange and cross-system transfer and between repository networks and with research information systems (CRIS)
Services	Share emerging practices for managing research data and linking data to publications Promote the role of repositories within research evaluation systems Develop a list of existing services deemed critical for the global research repository community (e.g. repository registries, publisher policy registries, etc.) and discuss how they could be better supported

V. Next Steps

The Aligning Repository Networks meeting in Rome marks the establishment of an 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.

Meeting delegates agreed that the ongoing work on aligning repositories should take place under the auspices of the Confederation of Open Access Repositories (COAR), an international alliance of repository initiatives. COAR has already been active in terms of promoting interoperability: a working group is currently developing an interoperability roadmap, and COAR has interest groups looking at usage data and controlled vocabularies. In addition, COAR already has significant international participation with members from 35 countries on 5 continents.

Aligning repository networks activities will extend the existing work being done in COAR. COAR will launch a Strategic Committee for Aligning Repository Networks, which will act as a forum for exchange of information between repository networks and identifying priorities. In addition, with input from the Strategic Committee, COAR will develop strategies for the implementation and take-up of priorities by the regional networks.

In many cases there are already bilateral discussions taking place. COAR will build on these existing activities and seek to further expand them to 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.

Appendix I

List of Participants

1. **Paul Ayris**, Director of University College London Library Services and President of LIBER (United Kingdom)

2. **Donatella Castelli**, Senior Researcher, CNR-ISTI, Italian National Research Center and Technical Coordinator, OpenAIRE (Italy)

3. **Carmen-Gloria Labbé**, Deputy General Manager, CLARA and LAReferencia and Vice Chairperson, COAR (Chile)

4. **Norbert Lossau**, Chairman of the Board, COAR and Vice-President, University of Göttingen (Germany)

5. **Rick Luce**, Co-Chair SHARE Steering Group and Associate Vice President for Research and Dean of University Libraries, University of Oklahoma (United States)

6. **Natalia Manola**, Department of Informatics & Telecommunications, University of Athens and Project Manager, OpenAIRE (Greece)

7. **Katharina Mueller**, Head of COAR Office, Confederation of Open Access Repositories (Germany)

8. **Brent Roe**, Executive Director, Canadian Association of Research Libraries (Canada)

9. **Eloy Rodrigues**, Director of the Documentation Services, University of Minho and Chair, COAR Repository Interoperability Working Group (Portugal)

10. **Kathleen Shearer**, Executive Director, Confederation of Open Access Repositories (Canada)

11. **Judy Stokker**, Deputy Vice-Chancellor, Technology, Information and Learning Support, Queensland University of Technology and Chair, Australian Open Access Support Group (Australia)

12. **Márta Virágos**, Deputy Director, University and National Library, University of Debrecen and Treasurer, COAR (Hungary)

13. **Tyler Walters**, Co-Chair SHARE Steering Group and Dean of University Libraries and Professor, Virginia Tech (United States)

14. **Xiaolin Zhang**, Executive Director, National Science Library of Chinese Academy of Sciences (China)

Appendix 2

Summary of Current State of Alignment

This is a non-comprehensive summary of the current state of alignment across repository networks. A more thorough review will be undertaken by COAR in the coming months.

Policies: Mandates and policies have not been implemented in all regions. The policies that are in place are usually implemented by individual funding agencies, although there some countries have passed open access (or repository) laws. Existing laws and policies vary significantly in their requirements across a number of elements as described in the table below.

Policy Elements	Current State
Method for open access: discipline-based repositories, institutional repositories, OA journals, or all.	• 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 a repository
Embargo periods	• 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 (Science, Technology, Engineering and Medicine) • Often require immediate deposit, with embargo access period
Content types	• 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	• Author's final manuscript is most common • Final published version

Re-use requirements	• 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)

Technical Interoperability	
Technical Issue	**Current State**
Metadata standards	• 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 RIOXX Guidelines. • In US, SHARE is developing metadata standards for the Notification System (currently assessing RIOXX). • NISO proposed Open Access Metadata and Indicators.
Usage statistics	• 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	• 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	Several initiatives exist: • 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)
Persistent identifiers Persistent identifiers for digital objects	Several initiatives exist: • 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.
Cross-system transfer Enabling multiple-deposit or transferring content from one system to another	• 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.

Services: There are a number of shared services that already exist with repository services.	
Service	**Description**
Publisher policies	Sherpa-Romeo: Publisher copyright & self-archiving policies
Discovery/ harvesting OAI-PMH is the basis for this approach, with a minimal number of additional fields layered on top of OAI-PMH.	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 BASE: Bielefeld Academic Search Engine collects, normalises, and indexes repository collections. BASE provides more than 50 million documents from more than 2,900 sources. You can access the full texts of about 75% of the indexed documents. The Index is continuously enhanced by integrating further OAI sources as well as local sources. 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. LAReferencia Portal: Harvests content from nine Latin American countries 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. National portals/harvesters
Repository directories Improves visibility of repositories	OpenDOAR: an authoritative directory of academic open access repositories. ROAR: Registry of Open Access Repositories
Cross-system transfer Enabling multiple-deposit or transferring content	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

from one system to another.	objects from one system to another.
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: "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)