# CURRENT STATE OF ALIGNMENT AND ROADMAP

OpenAIRE

Open Access Infrastructure for Research in Europe

December 9, 2016

OpenAIRE2020

Open Access Infrastructure for Research in Europe towards 2020
Deliverable Code: D3.3 - Version (1.0 – Final)
PUBLIC

This report provides an overview of the current state of repositories, the status of alignment across repository networks, and a vision and roadmap for next steps. In terms of current global alignment, the international repository community is diverse and at different stages of maturity and development across regions, countries and institutions. That being said, there has been significant progress in terms of aligning repository networks over the last two years as reflected in the activities reported here.


H2020-EINFRA-2014-1
Topic: e-Infrastructure for Open Access
Research & Innovation action
Grant Agreement 643410

# Document Description

## DX.X – Name of the deliverable

| | |
|---|---|
| WP3 - International Alignment | |
| WP participating organizations: COAR, CLARA, UMNIHO, UGOE UoA, CNR, | |
| Contractual Delivery Date: 10/2016 | Actual Delivery Date: 12/2016 |
| Nature: Report | Version: 0.2 (Final) |
| Public Deliverable | |

### Preparation Slip

| | Name | Organisation | Date |
|---|---|---|---|
| **From** | Kathleen Shearer | COAR | 9/12/2016 |
| **Edited by** | Alberto Cabezas, Tony Ross-Hellauer | CLARA, UGOE | 3/12/2016 |
| **Reviewed by** | Eloy Rodrigues | UMINHO | 7/12/2016 |
| **Approved by** | Eloy Rodrigues | UMINHO | 9/12/2016 |
| **For delivery** | Mike Chatzopoulos | UoA | 9/12/2016 |

### Document Change Record

| Issue | Item | Reason for Change | Author | Organization |
|---|---|---|---|---|
| V0.1 | Draft version | 1st draft | Kathleen Shearer | COAR |
| V0.2 | First Delivery | Review and edits to WP participants | Alberto Cabezas, Tony Ross-Hellauer, Eloy Rodrigues | CLARA, UGOE, UMINHO |
| V1.0 | Final version | Added end user feedback | Kathleen Shearer/Ilkay Holt | COAR |

# Table of Contents

# Disclaimer

This document contains description of the OpenAIRE2020 project findings, work and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium head for approval.

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of the OpenAIRE2020 consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 28 Member States of the Union. It is based on the European Communities and the member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors. (http://europa.eu.int/)


OpenAIRE2020is a project funded by the European Union (Grant Agreement No 643410).

# Acronyms

| | |
|---|---|
| BASE | Bielefeld Academic Search Engine |
| COAR | Confederation of Open Access Repositories |
| CORE | COnnecting REpositories |
| JAIRO | Japanese Institutional Repositories Online |
| Redalyc | Red de Revistas Cientificas de America Latina y el Caribe, España y Portugal |
| SciELO | Scientific Electronic Library Online |

# Publishable Summary

There is tremendous enthusiasm for openness globally as reflected in the momentum for open science, data sharing and open access, because it promises to greatly improve the impact and reuse of research results for social and economic benefit. As this trend continues and expands across the world, we need to ensure we are developing infrastructure and services that reflect the underlying principles of openness, while also supporting quality, equity and diversity in a truly global sense.

COAR's vision is to position institutions as the foundation for a distributed, globally networked infrastructure for scholarly communication, on top of which layers of value added services will be deployed, thereby transforming the system, making it more research-centric, open to and supportive of innovation, while also collectively managed by the scholarly community. In this scenario, research institutions collectively take on responsibility for preserving, curating and providing access to the valuable products created through research and scholarship, and connect together to create a global knowledge commons.

A nascent global network already exists through thousands of open access repositories and local journals that are operated by universities and research institutions around the world. Many of these initiatives are connected through national, regional and thematic networks, and COAR represents a mechanism for communication, defining standards and developing strategies across countries and regions.

This report provides an overview of the current state of repositories, the status of alignment across repositories and repository networks, and a vision and roadmap for next steps.

COAR's vision is to leverage the existing global repository network to position repositories and research institutions as the foundation for a distributed, globally networked infrastructure for scholarly communication, on top of which layers of value added services will be deployed, thereby transforming the system, making it more research-centric, open to and supportive of innovation, while also collectively managed by the scholarly community.

Many of the ideas and concepts presented in this report have originated from discussions from a meeting in Porto, Portugal in April 2015 and the COAR Next Generation Repositories Working Group, launched in April 2016. More information about these activities are available on the COAR website.

# 1 | THE MISSION OF REPOSITORIES

The mission of a repository is to manage and provide access to the valuable and diverse intellectual output of the community it serves. In this way, it offers a vital local service. Equally important, however, is that repositories are nodes in a larger network, contributing their collective contents to a global knowledge commons on top of which value added services can be built. By working together, repositories have the potential to offer a comprehensive view of the research of the whole world, while also enabling each scholar and institution to participate in the global network of scientific and scholarly enquiry.

Also, crucially, repositories represent a distributed and participatory model in which institutions manage content locally, but contribute to a global knowledge base through adoption of common, open standards. Open, distributed systems, such as the global network of repositories, have an inherent sustainability. They increase the resilience of infrastructure and foster social and institutional participation. Their strength lies in the absence of central control, which means they are not at risk of takeover or buy-out.

# 2 | REPOSITORY DEMOGRAHICS

Open access repositories play a variety of roles in the scholarly communication system, and these roles continue to expand and evolve. While much of the journal literature is still behind pay walls, repositories are mechanisms by which access to this literature is provided free of charge. As open access expands to the broader concept of open science, open access repositories will become indispensable for managing, tracking, and providing access to the full range of outputs produced through research.

Repositories began to appear on the scholarly landscape in the late 1990's, but their real growth in numbers has occurred over that last ten years, mainly because of the availability of open source repository platforms.

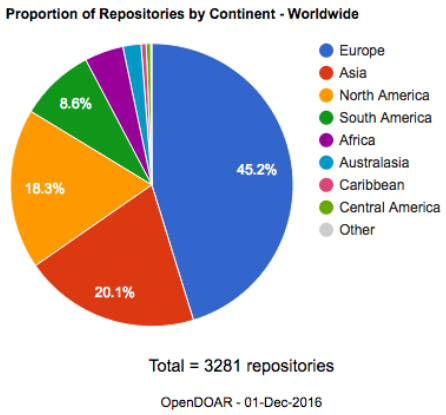


**FIGURE 1. PROPORTION OF REPOSITORIES BY CONTINENT**

As of December 1, 2016 there were 3281 repositories listed in OpenDOAR, a registry of repositories. This is 400 more than were listed 18 months ago. As reflected by OpenDOAR statistics, repositories are found in all regions. However, these figures likely under-represent the real number of repositories in some regions, especially outside of Western Europe and North America. For example, according to the National Institute of Informatics, there are currently over 700 repositories in Japan, which is a far higher than represented in OpenDOAR statistics.[1]

According to OpenDOAR, the vast majority of repositories, just under 85%, are institution-based and hosted mainly by research institutions or universities. There are also some very important and highly valued thematic repositories, including arXiv and PubMed Central, and journal networks, such as SciELO and Redalyc.

It is difficult to estimate the volume of 'collective' content in the world's repositories. The CORE (COnnecting REpositories) search service, run by Knowledge Media institute at the Open University, is one of the world's largest aggregators of full text content. CORE harvests over 37.5 million records and 4.5 million full text records. CORE estimates that, currently 35 million of its records come from repositories.[2] BASE, the Bielefeld Academic Search Engine operated by Bielefeld University Library is another large-scale global aggregator aggregates from a variety of data sources. According to BASE statistics, they have indexed 102,410,145 records from 6697 data providers, which represent both traditional repositories as well as journal platforms (99% of which are Open Journal Systems).[3]

[1] See JAIRO Cloud statistics: http://ju.nii.ac.jp/en/
[2] Figure comes from private communications with Petr Knoth, CORE Team and Product Lead
[3] Figure comes from private communications with Friedrich Summann, BASE Project Team

Repositories collect a wide variety of content types. The data from the OpenDOAR directory shows that most common content types contained in the repositories it indexes are journal articles, followed by theses and dissertations, and then books, and book chapters and sections. Additionally, the language of this content is quite varied. According to BASE statistics, 44% of their harvested documents are English, 31% the language is not indicated (there is no metadata about the language), and the other 25% is in other languages.[4]

[4] Figure comes from private communications with Friedrich Summann, BASE Project Team

# 3 | VALUE ADDED SERVICES FOR REPOSITORIES

A major value proposition for repositories lies in the potential to interconnect them to create a network providing unified access to research outputs that can be used by machines and human users. To a large degree, the types of services being developed on top of repository content will define the requirements for repositories to adopt common standards and behaviours.

Below is a list of the current, developing and potential services for repository networks:

**TABLE 1: REPOSITORY SERVICES**

| Level of Service | Service | Description | Standard Requirement |
|---|---|---|---|
| Current | Metadata harvesting | Metadata harvesting is the basic level service provided by the repository networks. It involved the automated process of collecting metadata or full-text records from different data providers to create useful aggregations of metadata and related services. | Adherence to the OAI-PMH, a common protocol for exposing metadata |
| Current | Discovery | The harvested metadata is used as the basis for search and discovery services. | Common vocabularies and metadata<br>Common approach to HTTP links with appropriate link types and format indicators on web resources that make up a scholarly object |
| Current | Tracking research outputs | This involves connecting research outputs with administrative information such as funders, projects and institutions. | Common vocabularies and metadata elements that indicate funder, project, and institution and resource type and availability |
| Developing | Text and data mining | This is the process of deriving information from machine-read material. It works by | Standardized approaches to describing re-use |

| | | | |
|---|---|---|---|
| | | copying large quantities of material, extracting the data, and recombining it to identify patterns.[5] | licenses |
| Developing | Routing, data syncing and notifications | Routers send notifications to repositories about research articles or other related content. Notifications may include articles themselves or information about new or modified metadata information. | Common protocols for interacting with repository networks |
| Potential | "Meta" science | Involves complex content analysis to better understand patterns of science and research, based in related content, co-authorship, clustering and so on. | More granular metadata and controlled vocabularies in order to identify other characteristics of objects |
| Potential | Usage measures | This will allow us to compare usage across papers and other content types. | Common methods of assessing usage of repositories, such as downloads, but also other indicators |
| Potential | Commenting, annotation and peer review | Repositories can add value by supporting commenting, annotation and peer review activities as functional layers on top of their collective contents. | Common approaches to supporting reviews, commenting and annotation |

[5] http://libereurope.eu/text-data-mining/

# 4| 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, funders and the public, regardless of location or disciplinary interest, to participate in and access research outputs worldwide. Aligning repository networks will foster the development of a more seamless scholarly infrastructure. It will also enable us to develop value added services, support the provision of uniform information about the impact of open access, and as well as avoid duplication of work across regions and enable cost synergies in areas of common interest.

While aligning practices is an important objective, COAR also recognizes the need for infrastructure and services to also support local and national requirements. There are different research priorities, technologies, and political and social challenges across the world. In addition, in a highly centralized ecosystem, a small number of actors can gain too much control and can establish a quasi-monopolistic position. Therefore, distribution of the control of scholarly resources is key principle that underlies our vision. Striking the right balance between local needs and a global network is important and we must find ways to realize these two fundamental conditions.

A major strategic priority for COAR is to align repository networks. To that end, COAR has been working with the regional and national aggregators, as well as representatives from other countries that do not yet have centralized services in Africa, Australia/New Zealand, Asia, Canada, and Latin America to promote and support greater alignment of repositories and repository networks. This activity, partially funded by the OpenAIRE2020 project, is being pursued through several avenues including bilateral agreements between networks, international working groups, strategic meetings, and informal discussions with regions and countries.

A key component of this strategy is to encourage data exchange across repository networks, so that there is redundancy across networks and multiple copies around the world. Underpinning this vision is that there is distribution of the control of scholarly resources (pre-prints, post-prints, research data, supporting software, etc.), so that no one actor can unduly influence for their benefit, rather than the benefit of the scholarly community.

The objectives of these discussions are as follows:

- To harmonize policies and practices to support the development of value added services across networks.
- To facilitate information exchange and shared problem solving around common challenges, services and issues.
- To increase the visibility of the various initiatives and demonstrate that networks can link research outputs to funders and projects across regions.
- To increase collaboration and avoid duplication of efforts by defining scope of services and distributing workload geographically.

TABLE 2. REGIONAL AND NATIONAL AGGREGATORS

| AGGREGATORS | DESCRIPTION |
|---|---|
| **Chinese IR Repository Grid** | In China, the situation is not as centralized as the other regional services. The Chinese Academy of Sciences manages the Chinese IR Grid, which harvests from 105 repositories that are managed by the institutes of the Chinese Academy of Sciences. Meanwhile, in the university sector, the CHAIR harvester (China Academy Institutional Repository), managed by CALIS, aggregates 85,732 records from 40 IRs. The National Academy of Sciences is currently considering how to combine these two services into one national harvester. |
| **JAIRO** | JAIRO is an abbreviation for Japanese Institutional Repositories Online. JAIRO is managed by the National Institute of Informatics. It collects metadata of institutional repositories in Japan. At the time of writing, JAIRO harvests from 571 data providers in Japan and aggregates 2,377,428 records. |
| **LA Referencia** | It is the network of open access repositories from nine Latin American countries (Argentina, Brazil, Ecuador, Chile, Columbia, Cost Rica, El Salvador, Mexico, and Peru). LA Referencia harvests scholarly articles and theses & dissertations from national nodes, which, in turn, harvest from repositories at universities and research institutions in each country. LA Referencia is based on agreement between the National Science Ministries and Science and RedCLARA. LA Referencia currently harvests from 9 national nodes and aggregates 1,392,945 records of open access content. |
| **OpenAIRE** | OpenAIRE is a European Commission funded project that supports the implementation of the EC Open Access Policy. OpenAIRE currently harvests from 6632 Journals, 541 Institutional Repositories and other data sources from around the world and contains 17,432,131 records. Its successor, OpenAIREplus, is aimed at linking the aggregated research publications to the accompanying research and project information, datasets and author information. |
| **SHARE** | SHARE is developing services to gather and freely share information about research and scholarly activities across their life cycle. The SHARE infrastructure was developed by the Association of Research Libraries in partnership with the Center for Open Science and SHARE is funded by the Institute of Museum and Library Services and the Alfred P. Sloan Foundation. SHARE is |

currently harvesting 4,287,471 records from 157 sources, mainly, but not exclusively, from US institutions.

Over the last several years, there has steady progress towards the harmonization of standards and adoption of common practices across repository networks.

**Recent activities with regions and countries outside of Europe**

1. **Africa**

- There have been various discussions about the feasibility of adopting an African harvester that could harvest all of the repositories in Africa, improving their visibility and supporting greater alignment and harmonization of repositories across the continent. One suggestion is that CODESRIA, the Council for the Development of Social Science Research in Africa, might be an appropriate organization to host a harvester. COAR is investigating this possibility of this, as well as the potential for external funding to support the development of the harvester.

**2. Australia and New Zealand**

- COAR has been working with several contacts in Australia and New Zealand to help the region assess the feasibility of adopting OpenAIRE metadata guidelines. Currently there is no centralized harvester in Australia or New Zealand, but they are still interested in harmonizing their metadata practices across the two countries and with the international community. They are currently in the process of assessing OpenAIRE Guidelines.

**3. Canada and United States**

- **SHARE (United States):** In July 2016, Kathleen Shearer, COAR Executive Director, attended the SHARE meeting. SHARE is a rapidly expanding network in the US. It was launched by ARL and other university associations, and it is managed by the Center of Open Science. COAR (with OpenAIRE and LA Referencia) has been working with SHARE to encourage the adoption of common standards, increase information sharing, and promote data exchange across networks. Because SHARE is harvesting from a wide range of data providers (beyond just repositories) they are reluctant to require their data provides to adhere to specific requirements.

- **Canada:** In November 2016, COAR co-organized a Next Generation Repositories Meeting in Ottawa, Canada with the Canadian Association of Research Libraries. The aim of this meeting was to bring together library directors and decision makers to develop a strategy for repositories in Canada. At this meeting, COAR presented the strategy for aligning repository networks and also our vision for next generation repositories.

**4. Latin America**

- **LA Referencia**: In September 2016, LA Referencia, the large regional network that harvest from 9 countries in Latin America (Argentina, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Mexico), agreed to adopt common guidelines with OpenAIRE (for more information, see section 3.3 under CLARA)

- **Chile:** OpenAIRE has started to develop a pilot with a local funder in Chile, CONECYT. The aim is to use OpenAIRE Service to support tracking CONECYT-funded research outputs.

**5. Asia**

- **China:** In September 2016, Kathleen Shearer attended the 4th annual Chinese Institutional Repository Conference took place in Chongqing, China on September 21-22, 2016, with the theme of "Rich functions and extended roles for institutional repositories". The conference was jointly organized by National Science Library, Chinese Academy of Sciences (CAS) and CIRG (Chinese institutional repository implementation group), and Chongqing University Library. It attracted over 330 registered participants, reflecting the growing interest in repositories in the Chinese library community. This further consolidated the relationship between COAR and the National Science Library at the National Academy of Sciences. At this meeting, the development of a national harvester for all of China was proposed. Since then, the National Science Library has indicated that it will develop a nation-wide repository harvester, "OpenSearch", to support visibility and other services on top of their repositories. The intention is that this harvester will also be a mechanism for greater international collaboration. COAR has been invited to provide expertise during these developments to ensure that the Chinese services are aligned with other international developments.

- **Japan:** OpenAIRE is harvesting the Japanese network of repositories through JAIRO, which is infrastructure managed by the National Institutes of Informatics. This has been achieved using crosswalks across the two diverse metadata schemas of JAIRO and OpenAIRE. JAIRO now represents the third largest data provider in OpenAIRE. NII is looking at adopting some of the OpenAIRE Guidelines and will consider which elements may be useful in the Japanese context.

- **Asia OA:** Asia OA is a special forum hosted by COAR in which members of the Asian open access community can share information, meet each other and build relationships. The community is dedicated to people working in the academic environment based in the Asian region. It celebrates Asian cultural diversity and unique way of doing things. Asia OA is a forum to help inform the of international developments and support the sharing of information in the Asian Community. In November 2016, OA Asia had a first meeting in Kuala Lumpur, Malaysia. There were representatives from China, Japan, Hong Kong, Singapore, India, Malaysia, Pakistan and Nepal. This meeting provided an opportunity to begin raising awareness with some of the less developed countries in Asia about the importance of national networks and international alignment.

**6. International**

- On April 12, 2016, COAR hosted an Aligning Repositories Networks meeting in Vienna, Austria. The aim of this meeting was to discuss strategic directions for repository networks in different region, and share challenges. The meeting was attended by representatives from Africa, Canada, China, Japan, Europe, Latin America, and the United States. One of the outcomes of the meeting was to develop a statement about the importance of supporting a variety of approaches to open access, and to consider various perspectives when adopting specific policies. The statement was co-published with UNESCO in May 2016 and has been widely disseminated around the world.

In terms of current global alignment, the international repository community is diverse and at different stages of maturity and development across regions, countries and institutions. Some countries do not have any central mechanism to help them harmonize their practices, and therefore practices can be extremely diverse. In other cases, there are communities of practice at varying levels of formal organization. The most mature jurisdictions are the countries with national and regional harvesters (or aggregators). These harvesters play a strong role in aligning repositories in their region by setting standards, supporting policy adherence, facilitating common practices and establishing collaborative communities. To that end, much of the efforts to align repository networks involve working with these most developed networks, as they have the authority, in many cases to define standards and practices. Currently there are regional/national aggregators in China, Europe, Japan, Latin America, and the United States, covering a significant portion of the world's repositories. That being said, there are also other means by which we can also work to improve alignment including working with national communities of practices around the world (often in the library community), the international aggregators (such as BASE and CORE), and the repository software platforms.

# 5 | NETWORK OBJECTIVES

The objectives of existing repository networks are already aligned to a large degree. Although the missions and objectives are articulated differently, the underlying aims of these networks are all in some way connected with visibility and access. In a more granular way, however, networks can have somewhat different emphases. Some are focused on tracking open access research, others on visibility, and others or enabling discovery and linking content.

TABLE 3. OBJECTIVES AS ARTICULATED BY THE DIFFERENT NETWORKS

| NETWORK | OBJECTIVE |
|---|---|
| **JAIRO** | JAIRO aims to strengthen the formation, securement, and information transmission of academic contents, NII, in close cooperation with academic institutions, such as universities, and assists formulation of institutional repositories in universities, etc. and cooperation among them. |
| **LA Referencia** | LA Referencia aims to create a truly comprehensive platform to support research and scholarly communication with the objective of increasing the visibility of science in participating countries. |
| **OpenAIRE** | OpenAIRE is an international socio-technical network that supports the implementation and monitoring of Open Science policies, including Open Access to publications and research data. |
| **SHARE** | SHARE is a higher education initiative whose mission is to maximize research impact by making a comprehensive inventory of research widely discoverable, accessible, and reusable. |

While most networks began with a specific focus on scholarly publications, especially journal articles, there has been a general broadening of scope to include a wider range of content such as research data. This reflects the growing interest by governments and the research community in open science. The networks are positioning themselves to support and integrate content types creating a more robust and integrated environment that enables integration and reuse of all scholarly output.

# 6| METADATA SCHEMAS AND VOCABULARIES

***Common metadata is the glue for interoperability***

As discussed earlier, many valuable services can only be built if there are common standards and functionalities applied to repositories. Given that most current services are built on metadata records, common metadata is particularly important. For example, to track publications related to a specific grant, funder or institution, the information about those entities needs to be included in a standardized way somewhere within the repository record. Although some networks, such as OpenAIRE and LA Referencia, have guidelines for metadata standards, there are still huge disparities in the way metadata is applied in individual repositories. This is true across all regions and countries. Even for jurisdictions with requirements, there remain challenges with metadata quality.

Both OpenAIRE and LA Referencia employ validators as a way of informing their data providers about levels of adherence with metadata guidelines. The validator validates both the optimal practices of the OAI-PMH transfer protocol and the correct use of metadata fields, providing a first estimation of the overall metadata quality and compatibility level. At least three of the regional networks also have services that improve the quality and comprehensiveness of the metadata they harvest once it has been aggregated. This is done through both automated and human mediated activities and is referred to as curation, cleaning or transformation.

In terms of aligning practices across networks, progress is being made towards greater harmonization. In September 2016, LA Referencia formally adopted the OpenAIRE Guidelines. This means that two of the largest repository networks are now aligned in terms of metadata requirements. In the future, LA Referencia will participate the development of the OpenAIRE guidelines to ensure they support the needs of Latin America as well as Europe. COAR has also been discussing with other regions the feasibility of further internationalizing the guidelines and assessing their relevance of beyond Europe and Latin America.

COAR has also been encouraging (meta)data exchange between repository networks as a way of building greater sustainability into the international repository system. This involves cross regional harvesting and enables harvesters to reflect a more global view of research outputs. Data exchange between repository networks will build redundancy across regions and ensure multiple copies are held around the world. Fundamental to this activity is a vision of distribution of control of scholarly resources. A bi-product of this activity is that it also fosters adoption of common metadata across the networks enabling more seamless aggregation. To date, OpenAIRE is harvesting metadata from both JAIRO and LA Referencia, and SHARE will soon begin to harvest from LA Referencia. Furthermore, there is already significant overlap in terms of harvesting repositories. For example, OpenAIRE and SHARE both harvest from the large thematic repositories, arXiv and PubMed Central.

Related to metadata schemas, is the issue of common vocabularies. Common vocabularies are an important aspect of interoperability because they ensure that everyone is using the same word to mean the same thing. Controlled vocabularies, if applied consistently, offer a common terminology for describing content, and greatly assist with the discovery, sharing and reuse of

content. COAR has a working group that has been developing a set of common vocabularies for specific metadata elements. In October 2016, COAR released the first vocabulary, "Resource Type" Controlled Vocabulary for Open Access Repositories, version 1.1. This vocabulary articulates the different types of content available in repositories (and other scholarly systems) in a standardized way. The vocabulary will is currently available in 12 languages – English, Catalan, Chinese, Dutch, French, German, Italian, Japanese, Portuguese, Russian, Spanish, Turkish. In the future, COAR will also produce several other multilingual controlled vocabularies including "date types", "version types", and "access modes". Although this work is progressing somewhat slowly, because it is being done on a voluntary basis, COAR will continue to pursue this activity as well as maintain and update the vocabularies in the future.

Although we have made significant progress with aligning repository networks and we are working closely with other the centralized services in other regions (Latin America, United States, China and Japan) there remains institutional challenges in terms of adoption of guidelines, standards and practices at the level of individual repositories. This is an issue that crosses all regions and countries, but is particularly acute in the "global south" where there are limited resources for travel, training and conference; as well as lower funding levels for operating the repository. These regions, especially, would benefit from greater training and capacity building activities and COAR is working with those regions to try to address these challenges.

# 7 | INTEGRATING REPOSITORIES WITH OTHER INFRASTRUCTURES

In addition to aligning repository networks, repositories must be integrated and/or interoperable with other systems with which they overlap, including research administrative systems (CRIS), research data repositories, journal publishing platforms, indexing and abstracting services and search engines.

This kind of integration is also happening via various forums including pilot projects at the institutional and network level, as well as increasing dialogue between the repository community and other important stakeholders. COAR has recently published a report, The "COAR Roadmap for Future Directions for Repository Interoperability" outlining priority areas for repository interoperability.

# 8 | NEXT GENERATION REPOSITORIES: COALESCING AROUND A COMMON VISION

The nearly ubiquitous deployment of repository systems in higher education and research institutions provides the foundation for a distributed, globally networked infrastructure for scholarly communication. However, repository platforms are still using technologies and protocols designed almost twenty years ago, before the boom of the Web and the dominance of Google, social networking, semantic web and ubiquitous mobile devices. This is, in large part, why repositories have not fully realized their potential and function mainly as passive, siloed recipients of the final versions of their users' conventionally published research outputs.

In order to leverage the value of the repository network, we need to equip it with a wider array of roles and functionalities, which can be enabled through new levels of web-centric interoperability. In April 2016, COAR launched a working group to identify the core functionalities for the next generation of repositories, as well as the architectures and technologies required to implement them; and to work with the repository community to help adopt these functionalities.

There are two threads to this work: (1) Increase the exposure by repositories of uniform behaviours that can be used by machine agents to fuel novel scholarly applications that reach beyond the scope of a single repository and that enable to smoothly embed repository content into mainstream web applications. (2) Integrate with existing scholarly infrastructures, specifically those aimed at identification, as a means to solidly embed repositories in the overall scholarly communication landscape. These threads will allow frictionless access to open content and encourage the creation of cross-repository added-value services.

The priority areas for this work are as follows:

- **Discovery**: Develop global interoperability of repositories through web-friendly repository technologies and architectures. The current reliance on OAI-PMH is not optimal for discovery and interoperability. Future repositories and repository networks should rely less on passive aggregation and develop services based on active dissemination such as peer-to-peer architectures, and subscription and notification-based information exchange.

- **Assessment**: Develop networked functionality related to the quality assessment of content. The aim is to define a model for overlay services on top of repositories using standardized registration, peer-review and quality assessment services thereby increasing the value of repositories significantly and expanding their role in the ecosystem. These types of services will generate symbolic value and provide researchers with the means to share their research, and be recognized and rewarded accordingly.

- **Workflows**: Expand the workflows and functionalities of repositories to better support the full lifecycle of research. Repositories need to be interoperable or integrated with other tools that authors are using such as Microsoft Word or Google docs, as well as with journal submission systems. Additionally, the repository should support more seamless deposit mechanisms, automatically recognize and auto-fill the metadata from a paper, dataset or other objects including author, title, date and so on into the submission form.

- **Use**: Define and adopt reliable and interoperable usage measures for repository content. Usage statistics and other indicators demonstrate to users the value of contributing content to repositories. The aim is to define standardized measures across repositories in order to establish trust and ensure reliable, community accepted measures.

To actualize this vision, we must fundamentally change the current functionalities of repositories, making them more web-friendly, and support the development of new, value added services on top.

In this scenario, alignment and interoperability of repositories will become even more critical. In order to support greater functionality at the network level, repositories will have to adopt common standards, technologies and behaviours. The COAR working group will share preliminary results with repository and scholarly communities in order to validate recommendations in December 2016. The aim is to have a final report, including recommendations, published in early 2017. These recommendations will be an important component of the aligning repository work in 2017 and on.

# 9 | COAR ROADMAP FOR ALIGNING REPOSITORY NETWORKS

This roadmap outlines the objectives and next steps to further align repository networks and support greater interoperability and the adoption of improved functionalities and services. In the coming months, COAR will expand on this roadmap by identifying the specific actions that will support progress across specific objectives.

1. **Increase the capacity and number of national and regional networks of repositories**

    1.1. Work with jurisdictions that do not have networks to support their adoption

    1.2. Promote and facilitate the transfer of existing networks and technologies into other regions to support greater progress and development of national networks

    1.3. Apply for funding to develop and undertake training for regional networks and repository managers

2. **Expand data exchange activities across repository networks**

    2.1. Work with existing regional and national networks to demonstrate the value of data exchange

3. **Expand the adoption of common vocabularies and metadata standards**

    3.1. Continue to develop and expand the COAR controlled vocabularies

    3.2. Promote the use of common vocabularies and metadata and raise awareness of the benefits in the repository community

    3.3. Further internationalize OpenAIRE Guidelines to support their adoption beyond Europe and Latin America by working with other stakeholders such as the international, regional and national networks, as well as national library associations to implement common metadata standards

    3.4. Encourage the development of plug-ins for vocabularies and metadata guidelines into the repository platforms

    3.5. Participate in discussions and activities with stakeholder communities (e.g. research administration, research data, digital heritage, etc.), to enable interoperability and common standards across systems.

4. **Coalesce around a common vision for next generation repositories**

    4.1. Complete and publish the work of the Next Generation Repositories Working Group

    4.2. Promote widely the vision and recommendations for the next generation of repositories

    4.3. Work with the repository platform developers to help implement recommendations

    4.4. Launch a pilot project for peer-review functionality on top of collective repository content

    4.5. Work with repository networks to implement other recommended functionalities, such as common usage measures

**5. Engage more deeply with other communities**

5.1. Increase engagement and work towards greater harmonization with domain repository communities, data repository communities, CRIS communities and open access publishers