

Roadmap for internationalization of metadata guidelines and vocabularies



SEPT. 2019

Advancing Open Scholarship

D5.2 – Roadmap for internationalization of metadata
guidelines and vocabularies

Version 1.0

PUBLIC

This deliverable presents a detailed plan for aligning with
global participants: stakeholders to involve, timeline, etc.



H2020-EINFRA-2017
Grant Agreement 777541

Document Description

D5.2 – Roadmap for internationalization of metadata guidelines and vocabularies

WP5 - Global OA Scholarly Communication Infrastructure	
WP participating organizations: COAR, CLARA, UOA, UGOE, CNR, UNIWARSAW, UMINHO, EIFL	
Contractual Delivery Date: 03/2019	Actual Delivery Date: 09/2019
Nature: Report	Version: 1.0 Final
Public Deliverable	

Preparation Slip

	Name	Organisation	Date
From	Kathleen Shearer Ilkay Holt Paul Walk	COAR	08/2019
Edited by	Kathleen Shearer	COAR	15/09/2019
Reviewed by	Alberto Cabezas Natalia Manola Jochen Schirrwagen	CLARA UOA UNIBI	15/09/2019
Approved by	Natalia Manola	UoA	17/09/2019
For delivery	Mike Chatzopoulos	UoA	19/09/2019

Revision History

Issue	Item	Reason for Change	Author	Organization
V0.1	Draft version	Integrate community feedback from May workshop	Kathleen Shearer	COAR
V0.2	First Delivery	Revision for more detailed information	Kathleen Shearer	COAR
V1.0	Final version	Revision based on 2nd round of reviews	Kathleen Shearer	UGOE

Table of Contents

1	Introduction	6
2	Current Context	7
3	Roadmap	9

Disclaimer

This document contains description of the OpenAIRE-Advance 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 OpenAIRE-Advance consortium and can in no way be taken to reflect the views of the European Union.

OpenAIRE-Advance is a project funded by the European Union (Grant Agreement No 777541).



Acronyms

API	Application Programming Interface
BASE	Bielefeld Academic Search Engine
CERIF	Common European Research Information Format
COAR	Confederation of Open Access Repositories
OAI-PMH	Open Archives Initiative - Protocol for Metadata Harvesting
RDA	Research Data Alliance

Publishable Summary

Aggregating metadata records from repositories and other data providers in order to provide regional and global discovery and analytics services requires metadata normalization. The quality, consistency and interoperability of metadata in repositories, therefore, is an important aspect that can influence the quality of these services and ultimately, the value of repositories in general. Repositories follow consistent guidelines, protocols, and standards for interoperability such as OpenAIRE Guidelines. The guidelines set important requirements with regards to the integration into repository platforms and have gained geographical significance, on the other hand, the current process for contributing to the development of guidelines is not representative of the expanding user community, nor is it well understood to those outside of OpenAIRE. There is an opportunity to formalize the approach and governance of the OpenAIRE guidelines to reflect their evolution.

Based on this context, COAR is proposing to undertake several activities that will contribute to improving the quality, comprehensiveness and alignment of metadata in repositories internationally. While there are other efforts to improve scholarly metadata, such as metadata2020 and the RDA Metadata Directory, the activities outlined in this roadmap will specifically target the repository community by working with regional networks and other regional associations, who have more direct access with the local organizations managing repositories. This document outlines the rationale and roadmap that will help advance alignment of repository metadata and begin to define a more representative governance structure for the OpenAIRE guidelines that are starting to be adopted outside of Europe.

1 | INTRODUCTION

Aggregator services collect and combine metadata (and sometimes full text) from repositories and other data providers and expose the aggregated contents in different interfaces (end-user orientated, OAI-PMH and other APIs, as well as downloadable data collections). These services, such as OpenAIRE, LA Referencia, BASE, CORE and others, provide regional and global discovery and analytics services for users by aggregating metadata records from repositories and other data providers. A certain level of metadata normalization is needed in order to integrate the different records and deliver services of value to users. The quality, consistency and interoperability of metadata in repositories, therefore, is an important aspect that can influence the quality of these services and ultimately, the value of repositories in general.

The potential to create a unified body of scholarly materials is reliant on interoperability – specifically, that repositories follow consistent guidelines, protocols, and standards for interoperability which allow them to communicate with each other and with other systems in order to transfer information, metadata, and digital objects. Although automation will enable us to reduce our reliance of some types of metadata (also as we expand our use of identifiers), there will still be the need for the adoption of some common vocabularies and metadata, especially to support transactional metadata between the networks and the resources.

2 | CURRENT CONTEXT

The [OpenAIRE Guidelines](#) are being actively promoted and used in Europe and Latin America (as well as to some extent in other regions). There are also a number of other national guidelines being adopted elsewhere, including RIOXX in the UK, JPCOAR in Japan, DINI-certificate in Germany and recent guidelines developed in the African context, DataCite metadata for research data, CERIF for research information and many domain based schemas. However, for publication and many other types of repositories, the default metadata of simple Dublin Core expressed through OAI-PMH is likely the most common de facto standard used in repositories currently.

In 2014-2015, COAR, along with CASRAI and several regional networks, undertook an effort to standardize some metadata elements across the networks and improve the quality of metadata adoption in repositories internationally. The aim was to develop a small number of COAR-CASRAI profiles which include underlying agreement on metadata elements, their associated definitions, and the vocabulary elements. The outcome of this work was an analysis of different practices across participating networks and a proposed strategy for harmonizing metadata elements. However, due to other priorities of the networks, there were no substantial local efforts to implement this strategy after it was published.

At a workshop organized by COAR in May 2019, COAR again raised the issue of metadata harmonization including alignment around a common set of metadata guidelines that could be adopted across regions and countries. Participants discussed the significant challenges related to this objective because the use cases for each network are different for different services and in different regions, resulting in the need for unique metadata elements in different contexts. In addition, in a world of increasing complexity and abundance of metadata, an approach based on static, published mappings may not succeed.

In conclusion, there is little interest in COAR developing a new, global metadata format. However, there is widespread acknowledgement that the community needs to address the issue of low quality and comprehensiveness of metadata currently and there is support for COAR to raise awareness of the importance of metadata and vocabularies. COAR already maintains several vocabularies that can be integrated into existing metadata schemas. Controlled vocabularies ensure that “everyone is using the same word to mean the same thing” and are key to achieving the COAR vision of a global knowledge commons, based on an interoperable, international network of open repositories. The three vocabularies [Resource Type Vocabulary](#), [Access Rights Vocabulary](#) and [Version Type Vocabulary](#), are openly available for use in all repositories and available in multiple languages. In terms of metadata, it is proposed that COAR could play a role in linking to the various different metadata schemas in use around the world, which could be useful for people choosing metadata in their repositories, and promoting the need for quality metadata in repositories.

The OpenAIRE Guidelines are built on the DRIVER-Guidelines and extend them to track access rights and funding information. Finally they have been "specialized" into Guidelines for institutional repositories, data repositories, software repositories and CRIS. However, as

OpenAIRE expands its reach, harvests more repositories, and enters into data exchange agreements with other networks, the OpenAIRE Guidelines are becoming more widely adopted both within and outside of Europe. For example, LA Referencia, a network of 10 countries in Latin America, adopted the guidelines [version 3](#) in 2015 and is moving towards the adoption of [version 4](#). In addition, the aims and objectives of OpenAIRE are expanding and include other types of services such as the broker, common usage statistics, etc. These new services may lead to new metadata requirements. While, on the one hand, the guidelines have set important milestones with regard to the integration into repository platforms and have gained geographical significance, on the other hand, the current process for contributing to the development of guidelines is not representative of the expanding user community, nor is it well understood to those outside of OpenAIRE. There is an opportunity to formalize the approach and governance of the OpenAIRE guidelines to reflect their evolution.

3 | ROADMAP

Based on the context described above, COAR proposes to undertake several activities that will contribute to improving the quality, comprehensiveness and alignment of metadata in repositories internationally. While there are other efforts to improve scholarly metadata, such as [metadata2020](#) and the [RDA Metadata Directory](#), the activities outlined in this roadmap will specifically target the repository community by working with regional networks and other regional associations, who have more direct access with the local organizations managing repositories.

Objectives:

- To provide a tool that compares of meta(data) schemas or guidelines commonly used in the repository community. This resource will be complemented with documented use cases, commonly used guidelines, and materials for advocacy and training purposes.
- Define a process for extending the governance and development process for the OpenAIRE guidelines as they are adopted beyond Europe in the future.

Scope:

Metadata schemas or application profiles which are in scope of the tool will tend to have the following characteristics:

- They are used in more than one repository
- They have broad applicability: domain-specific metadata schemas and profiles may be in scope, but only where they are widely used
- They are openly documented

Benefits:

- Identify and reuse published metadata schemas
- Implement metadata schemas for the purpose of interoperability
- Construct mappings based on others' published schemas
- Understand better the need for quality metadata
- Increase value of their repositories

Five Interactive Activities:

The roadmap involves five iterative activities (described in more detail on the following page):

1. Identify different metadata schemas in use around the world, including information about the use case and major differences
2. Design a basic tool for listing and comparing the different schemas
3. Provide access to the existing resources to assist repository managers in implementing better metadata guidelines in their repositories, and develop resources where none exist
4. Undertake an awareness campaign about the need and value of adoption of quality metadata and vocabularies in repositories
5. Define a governance framework for the OpenAIRE guidelines to include other stakeholder communities in the process of updating guidelines

Action	Description	Date
Identification of different metadata schemas currently in use	This will involve an international scan of metadata schemas in use in the context of repositories. The scope of this will include information about the different use cases and identify similarities and differences. In many cases, it is expected that the metadata being assigned to repository records vary according to submitting author and other factors. A short survey will be developed to support information gathering and launch the awareness raising campaign.	Sept - Oct 2019
Design a tool for comparison of relevant schemas	A tool on the COAR website will be developed that provides links to the existing metadata schemas in use internationally.	Nov 2019
Provide access to the existing resources to assist repository managers in implementing better metadata guidelines in their repositories, and	The metadata schemas webpage will be enhanced with links to other resources, including good practice guidelines and other available information to help improve the quality of the metadata in repositories. In addition, where there are	Nov - Dec 2019

<p>develop resources where none exist</p>	<p>obvious gaps in existing resources, COAR will identify gaps, and convene the community in order to develop new resources that are needed. At the same time, COAR will develop other resources more appropriate for raising awareness of the issue, that can be used in the awareness campaign launched in January 2020.</p>	
<p>Undertake an awareness campaign about the need and value of adoption of quality metadata and vocabularies in repositories</p>	<p>This campaign will articulate the value and importance of quality metadata for the repository community, and encourage the adoption of existing schemas and standards. The main target of the campaign will be repositories and institutions (libraries, research centres, etc.) that maintain repositories, however, other communities (researchers, journals and funders) will also be targeted in some of the messaging including and therefore will liaise with other related initiatives like metadata2020 to coordinate messaging. The campaign will be delivered through social media, mailing lists, webinars and posters presented at relevant conferences.</p>	<p>Jan - Mar 2020</p>
<p>Define a process and governance framework for OpenAIRE guidelines</p>	<p>A working group from the major user groups of the OpenAIRE Guidelines will be convened to define a process and a governance structure for guidelines development in the future. The participants in the working group will be convened by COAR to propose a more inclusive process for the next version of the guidelines, which will include aspects such as defining major use cases, mapping use cases to metadata elements, and timelines. The working group will also define light weight governance structure to ensure that the</p>	<p>Jan - Mar 2020</p>

	next version of the guidelines reflect the expanding user community.	
--	--	--