OpenAIRE Guidelines for
Data Archive, Literature Repository and CRIS Managers

Pedro Príncipe
University of Minho, Portugal
pedroprincipe@sdum.uminho.pt

Eloy Rodrigues
University of Minho, Portugal
eloyrodrigues@sdum.uminho.pt

Najla Rettberg
University of Goettingen, Germany
najla.rettberg@sub.uni-goettingen.de

Jochen Schirrwagen
University of Bielefeld, Germany
jochen.schirrwagen@uni-bielefeld.de

Mathias Loesch
University of Bielefeld, Germany
mathias.loesch@uni-bielefeld.de

Mikael Karstensen
Elbæk Technical University of Denmark, Denmark
miel@dtic.dtu.dk

Lars Holm Nielsen
CERN, Switzerland
lars.holm.nielsen@cern.ch

Keywords: Openaire, repositories, research data, open access.

1. OpenAIRE–Open Access Infrastructure for Research in Europe

OpenAIRE supports the European Commission Open Access policy by providing an infrastructure for researchers to comply with the European Union Open Access mandate. The OpenAIREplus project – 2nd Generation of Open Access Infrastructure for Research in Europe – is a 30 month project funded by the European Commission 7th Framework Programme, and extends the mission of OpenAIRE (initiated in December 2009) further to facilitate access to the entire Open Access scientific production of the European Research Area, providing cross-links from publications to data and funding schemes. This large-scale project brings together 41 pan-European partners, including three cross-disciplinary research communities. The project will capitalize on the successful efforts of OpenAIRE, which is rapidly moving from implementing the European Union Open Access Pilot project into a service phase, enabling researchers to deposit their 7th Framework Programme and European Research Area funded research publications into Open Access repositories.

Exposure and visibility of content from a range of European repositories will be significantly increased when a common and interoperable approach is taken and care to adhere to existing guidelines. This compatibility will lead to future interoperability between research infrastructures, and structured metadata is of benefit to individual data repositories and the knowledge community at large.

OpenAIRE is starting to move from a publication infrastructure to a more comprehensive infrastructure that covers all types of scientific output. To put this into practice an integrated suite of guidelines were developed with specific requirements supporting the goal of OpenAIRE and the European Commission.

By implementing all three sets of the OpenAIRE Guidelines, repository managers will be able to enable authors who deposit publications in their repository to fulfill the EC Open Access requirements, as well as the requirements of other (national or international) funders with whom OpenAIRE cooperates. In addition it will allow the OpenAIRE infrastructure to add value-added services such as discoverability and linking, and creation of enhanced publications. In short, building the stepping-stones for a linked data infrastructure for research.

The poster briefly outline the OpenAIRE Guidelines, highlighting the set of guidelines for Data Archive Managers, for Literature Repository Managers and for CRIS Managers.

2. OpenAIREplus: Supporting Interoperability through Guidelines

2.1. Guidelines for Literature Repository Managers

The OpenAIRE Guidelines for Literature Repository Managers provide orientation for repository managers to define and implement their local data management policies according to the requirements of OpenAIRE.

Initially, the requirements of the OpenAIRE infrastructure were established to support and monitor the implementation of the EC open access pilot. OpenAIRE is now widening its scope, both in terms of infrastructure and content. In this context, the current Guidelines for Literature Repository Managers (version 3.0) are intended to guide repository managers to expose content not only EC funded publications, but also other Open Access publications, regardless of their funding.

By implementing these Guidelines, repository managers will be able to enable authors who deposit publications in their repository to fulfill the EC Open Access requirements, as well as the requirements of other (national or international) funders with whom OpenAIRE cooperates. In addition it will allow the OpenAIRE infrastructure to add value-added services such as discoverability and utilization. According to the ongoing expansion, the new version of guidelines anticipates the merger of the DRIVER Guidelines into the context of OpenAIRE Guidelines.

OpenAIRE uses the OAI-PMH for harvesting publication metadata and OpenAIRE expects metadata to be encoded in the Dublin Core metadata format (metadata-Prefix oai_dc). OpenAIRE relies on a specific syntax used in the values of standard Dublin Core metadata fields to identify projects, funders, referenced publications, and datasets. This syntax takes the form of URIs and is defined as the info:eu-repo namespace

2.2. Guidelines for Data Archive Managers

The OpenAIRE Guidelines for Data Archive Managers provides instruction for data archive managers to expose their metadata in a way that is compatible with the OpenAIRE infrastructure. Data archives should be included in the OpenAIRE information space when data are related to a document e.g. a dataset cited by an article. These guidelines are largely based on the DataCite metadata schema.

OpenAIRE has adopted the DataCite Metadata Schema as the basis for harvesting and importing metadata about datasets from data archives. The core mission of DataCite is to build and maintain a sustainable framework that makes it possible to cite data through the use of persistent identifiers, DOIs.

OpenAIRE believes that this will provide a domain agnostic metadata schema and provide interoperability through a small number of properties.

By implementing the OpenAIRE Guidelines data archive managers are facilitating the creation of enhanced publications and building the stepping-stones for a linked data infrastructure for research.

2.3. Guidelines for CRIS Managers

OpenAIRE and euroCRIS have agreed on the use of CERIF-XML as the standard format by which OpenAIRE will harvest information from CRIS. The Guidelines for CRIS Managers have provided a specification of the application profile based on the requirements of the OpenAIRE data model, relevant vocabularies, data entities and description and their values. This is a step forward in the collaboration of the research information and the repository communities, and the use of CERIF-XML as the basis for an exchange format within the research information domain.

3. Guidelines wiki

In an effort to make these guidelines as open as possible to the wider public, OpenAIREplus project have established a Guidelines Wiki—http://guidelines.openaire.eu. The intention of this wiki is to provide a public space to share OpenAIREs work on interoperability and to engage with the community.

This wiki is open and intended for all stakeholders and data providers to add their experiences and best practices. The wiki was publicly announced last June 2013 and the community is welcome to participate, post experiences, add comments, and to promote to all repository managers (publications, CRIS, data).

The OpenAIRE Guidelines wiki have all the needed information about the three sets of Guidelines: 1) for literature repository managers, 2) for data archive managers, 3) for CRIS Managers (currently in draft).

Additionally to the wiki the OpenAIRE services provides an easy way to validate both publication and data repositories guidelines at http://validator.openaire.eu. Other support information (Helpdesk, FAQ, Guides) is available in the OpenAIRE portal—http://www.openaire.eu.