OpenAIRE Guidelines: Supporting Interoperability for Literature Repositories, Data Archives and CRIS

Príncipe, Pedro; Rettberg, Najla; Rodrigues, Eloy; Elbæk, Mikael Karstensen; Schirrwagen, Jochen; Houssos, Nikos; Nielsen, Lars Holm; Jörg, Brigitte

Published in:
Procedia Computer Science

Link to article, DOI:
10.1016/j.procs.2014.06.015

Publication date:
2014

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2014) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

CRIS 2014

OpenAIRE Guidelines: supporting interoperability for Literature Repositories, Data Archives and CRIS

Pedro Príncipe[a], Najla Rettberg[b], Eloy Rodrigues[c], Mikael K. Elbæk[d], Jochen Schirrwagen[e], Nikos Houssos[f], Lars Holm Nielsen[g], Brigitte Jörg[h]

[a]*University of Minho, Portugal, pedroprincipe@sdum.uminho.pt*
[b]*University of Goettingen, Germany, najla.rettberg@sub.uni-goettingen.de*
[c]*University of Minho, Portugal, eloyrodrigues@sdum.uminho.pt*
[d]*Technical Information Center of Denmark, Technical University of Denmark, Denmark, miel@dtic.dtu.dk*
[e]*University of Bielefeld, Germany, jochen.schirrwagen@uni-bielefeld.de*
[f]*National Documentation Centre / National Hellenic Research Foundation, Greece, nhoussos@ekt.gr*
[g]*CERN, Switzerland, lars.holm.nielsen@cern.ch*
[h]*JeiBee Ltd., United Kingdom, brigitte.joerg@gmail.com*

Abstract

OpenAIRE – Open Access Infrastructure for Research in Europe – is moving 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. This poster outlines the OpenAIRE Guidelines, highlighting the set of guidelines for Literature Repository Managers, for Data Archive Managers and for CRIS Managers.

Peer-review under responsibility of euroCRIS.

Keywords: OpenAIRE, repositories, research data, CRIS systems, interoperability, 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

Peer-review under responsibility of euroCRIS.

publications to data and funding schemes. This large-scale project brings together 41 pan-European partners, including three cross-disciplinary research communities. The project capitalizes 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, data archives and CRIS 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 moving 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 following OpenAIRE Guidelines, managers of scholarly communication systems across Europe will be able to support authors 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.

This poster outlines the OpenAIRE Guidelines, highlighting the set of guidelines for Literature Repository Managers, for Data Archive 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 not only EC funded publications, but also other Open Access publications, regardless of their funding.

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.

According to the ongoing expansion, the new version of guidelines anticipates the merger of the DRIVER Guidelines – http://www.driver-support.eu/managers.html – into the context of OpenAIRE Guidelines.

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

CERIF-XML has been agreed as the standard format by which OpenAIRE will harvest information from CRIS systems. Information retrieval from individual CRIS systems by OpenAIRE is an example of a point-to-point data exchange among CRIS systems, since OpenAIRE itself can be considered a CRIS system; therefore, CERIF XML is the obvious choice as an interoperation standard. Import of information from CRIS systems is facilitated by the fact that the data model of the OpenAIRE 2nd generation infrastructure is CERIF-compliant [1].

The Guidelines provide orientation for CRIS managers to expose their metadata in a way that is compatible with the OpenAIRE infrastructure. By implementing the Guidelines, CRIS managers support the inclusion and therefore the reuse of metadata in their systems within OpenAIRE.

CERIF is a comprehensive domain model, having a wider scope than the OpenAIRE information space, which currently does not aim to represent the full range of information in CERIF CRIS system. Thus, the Guidelines essentially specify an OpenAIRE-specific subset of the CERIF model in terms of data elements and also the particular OpenAIRE semantics. The semantics have the form of classifications and classifications schemes (terms and vocabularies) of the CERIF Semantic Layer, which is the approach utilised within the OpenAIRE data model to represent classifications and relationships semantics. Harvesting of information from CRIS systems will be performed using the OAI-PMH protocol.

2.4. 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:

- Literature Repositories: https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_Literature_repositories,
- Data Archives: https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_Data_Archives,
- CRIS: https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_CRIS.

Additionally to the wiki the OpenAIRE services provides an easy way to validate both publication and data repositories guidelines at http://validator.openaire.eu. Support for CRIS guidelines validation is planned for 2014. Other support information (Helpdesk, FAQ, Guides) is available in the OpenAIRE portal – http://www.openaire.eu.

Acknowledgements

The work presented in this paper has been developed under the OpenAIREplus Project (Ref No: 283595) of the EU-funded FP7- INFRASTRUCTURES Programme.

References

1. Manghi, P., et al. (2012). The Data Model of the OpenAIRE Scientific Communication e-Infrastructure. Metadata and Semantics Research, Communications in Computer and Information Science, Springer, Vol. 343, pp. 168-180.