The OpenAIRE Data Infrastructure Services: On Interlinking European Institutional Repositories, Dataset Archives, and CRIS Systems

Paolo Manghi[1] , Lukasz Bolikowski[2], Natalia Manola[3],
Jochen Schirrwagen[4], Najla Rettberg[5], Tim Smith[6]

[1]ISTI-CNR, Italy, paolo.manghi@isti.cnr.it
[2]ICM, Poland, l.bolikowski@icm.edu.pl
[3]NKUA, Greece, natalia@di.uoa.gr
[4]Bielefeld University Library, Germany, jochen.schirrwagen@uni-bielefeld.de
[5]University of Goettingen, najla.rettberg@sub.uni-goettingen.de
[6]CERN, Switzerland, tim.smith@cern.ch

Abstract

The OpenAIREplus project (Open Access Infrastructure for Research in Europe Plus, Dec 2011 – May 2014) aims at realizing and maintaining the European scholarly communication data infrastructure.[1] To this aim, the data infrastructure collects and interlinks content from repositories (i.e., OA and non-OA publications), dataset archives (i.e., research datasets), and CRIS systems (i.e., funding information from European Commission and National funding schemes), and supports advanced metrics to measure impacts of Open Access mandates and funding over research. The infrastructure populates a graph of (metadata of) objects spanning across all research disciplines and countries, with the major objectives of: (*i*) providing enhanced access to the graph for end-users and third-party systems, (*ii*) experimenting automatic inference of semantic relationships between different object typologies (e.g., datasets and publications), and (*iii*) construction and refinement of "enhanced publications". This poster presents the high-level architecture, data model, and functionalities of the data infrastructure services devised in the context of the OpenAIREplus project. Services provide functionalities for *data collection and integration*, i.e., registering-validating data sources to the infrastructure in order to collect and convert their content into OpenAIREplus objects; *data inference*, i.e., analysis of collected metadata and digital objects (e.g., PDFs), to automatically infer new objects or relationships between them (e.g., content similarity, object citation, content classification); *data insertion and curation*, i.e., addition, removal or update of collected objects by a set of experts or registered end-users; *data provision*, i.e., standard APIs and interfaces to search, browse and visualize the graph of objects.

---

[1] OpenAIRE web site, http://www.openaire.eu