Repository power: how repositories can support Open Access mandates

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

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

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

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

Paolo Manghi, Consiglio Nazionale delle Ricerche, Italy, paolo.manghi@isti.cnr.it

Natalia Manola, National Kapodistrian University of Athens, Greece, natalia@di.uoa.gr

José Carvalho, University of Minho, Portugal, jcarvalho@sdum.uminho.pt

Session Type (select one)

☐ Panel
X Presentation

Abstract

Many funding agencies have Open Access mandates in place, but how often are scientific publications as outputs linked to funding details? The benefits of linking funding information to publications as part of the deposit workflow can assist in adhering to Open Access mandates. This paper examines how OpenAIRE – Open Access Infrastructure for Research in Europe – can ease monitoring Open Access and reporting processes for funders, and presents some results and opportunities. It also outlines how it relies on cleaned and curated repository content, a vital cog in the ever turning wheel of the global scholarly landscape, and the benefits it brings.

Conference Themes

Select the conference theme(s) your proposal best addresses:

X Supporting Open Scholarship, Open Science, and Cultural Heritage
☐ Managing Research (and Open) Data
☐ Integrating with External Systems
☐ Re-using Repository Content
☐ Exploring Metrics and Assessment
☐ Managing Rights
☐ Developing and Training Staff
☐ Building the Perfect Repository

Keywords

OpenAIRE, Open Access, Repositories, Funding Information.

Audience

Repository managers, Developers, Funding Agencies, Research Managers, Data producers, Librarians.

Introduction

Until now, funders and research institutions, who are increasingly concerned with tracking their outputs in order to consolidate and assess results, have no precise way of knowing exactly which and how many research publications and research data result from the resources they allocate to research. Many funders have now stipulated open access to results, and many have OA mandates in place, but how often are these outputs linked to funding details? The benefits of linking funding information to publications as part of the deposit workflow can assist in adhering to OA mandates. This presentation will examine how OpenAIRE[i], a European connected repository infrastructure can ease monitoring open access and reporting processes for funders. It also outlines how it relies on cleaned and curated repository content, a vital cog in the ever turning wheel of the global scholarly landscape, and the benefits its brings.

Background

The European Commission (EC) has run a pilot on OA in FP7 in seven research areas to publications for the past five years, and with the new Horizon2020 funding programme[ii], OA is now strictly mandated for all research areas: all publications must be deposited in an open access repository (institutional or disciplinary) and made openly accessible immediately or no longer than 6 or 12 months (for SSH)[iii]. OpenAIRE supports this EC policy, by having developed a repository-centric infrastructure. Launched in January 2015, OpenAIRE2020 now encourages all European repositories to participate in this infrastructure, so that EC-funded research results can be made openly visible. OpenAIRE is one of many complementary initiatives (e.g. FundRef, SHARE, CHORUS, Jisc-RIOXX) which are working towards the identification, tagging and linking of funders and projects with publications. OpenAIRE also works to serve other stakeholders such as project coordinators, research administrators, researchers who wish to link their publications to data, and providing services to national funders.

Implementation

The design of the OpenAIRE infrastructure is such that it relies on the participation of existing repositories (and other data sources). In order to facilitate that participation, repository managers are supported by a set of OpenAIRE Guidelines to align their local data management policies with the requirements of the OpenAIRE infrastructure[iv]. They provide guidance on how to interpret and expose metadata and how to get validated and aggregated into the OpenAIRE

infrastructure. The guidelines have adopted established and existing practices with different classes of content providers: i) guidelines for Literature Repositories using Dublin Core, ii) guidelines for Data Repositories using Datacite Metadata Schema, iii) guidelines for CRIS systems based on CERIF-XML. They have in common the recommended use of controlled vocabularies (for publication type, language and date encoding, rights information and access mode); of persistent identifiers for publications, datasets; of proper information encoding about funding bodies and project grantIds.

OpenAIRE strongly encourages reference to funding at time of deposit. Repositories and journals should extend their metadata description to accommodate funding information and expose in a standard format. The OpenAIRE guidelines have gone some way to make uniform these descriptions. Additionally, Repositories must be able to ingest basic and public authoritative funder information (i.e., funder name and project id) into their databases. In this way, researchers either (i) select from a list or (ii) systems are able to automatically identify funding information through text mining the scientific publications.
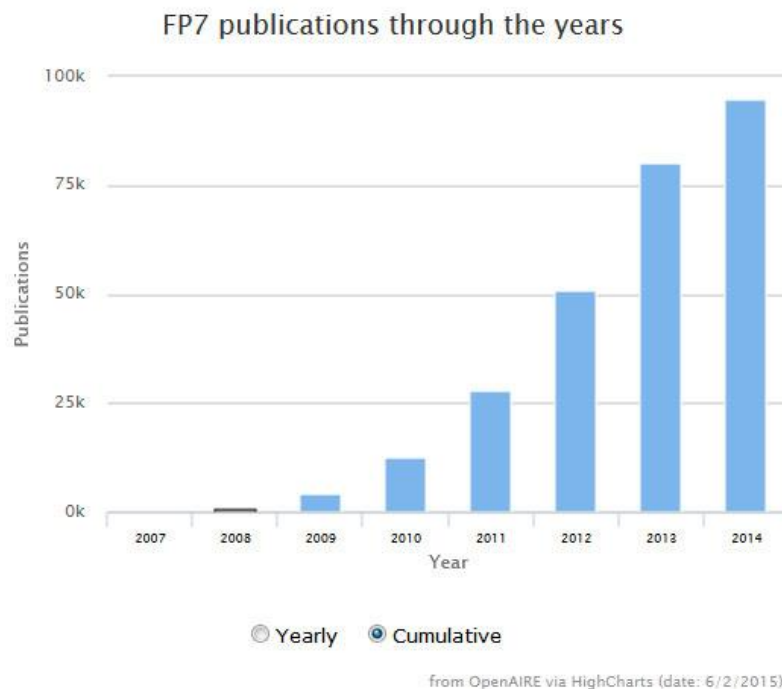

Figure 1 - Growth of publications linked to FP7 projects: 95,427 publications in 10,463 FP7 projects (from a total of 24,691 FP7 projects); 49,689 publications are OA, 2,840 are restricted and 162 are still in embargo

Challenges

There are a number of obstacles OpenAIRE has faced along the way, for example, getting buy-in from repository community for uptake of the OpenAIRE guidelines, and getting funding
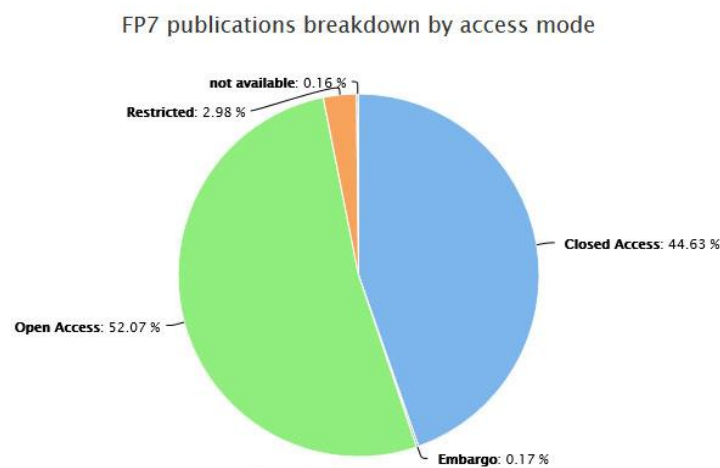
information into the repository workflow. General metadata quality in some repositories can be poor, and OpenAIRE has to do much curation to enhance records once they are harvested. In addition, strong communication from a funder to project coordinators about how to implement the OA mandate is vital to the uptake of such a mandate. Another challenge is requesting the relevant information from funders: funding organizations must expose the basic and public information (i.e., funder name/funding scheme/project id) of their funding programmes to repositories. The OpenAIRE guidelines will continue to be developed by OpenAIRE2020, a 42 month project (2015 - 2018) funded by the EC's Horizon2020 Programme. The main objective is to establish an open and sustainable scholarly communication infrastructure.

Opportunities

This paper will also present some results and opportunities that adding funding information to repositories can lead to. At present there are 580 participating data sources (compliant repositories, ejournals, aggregators and data repositories). In addition to funding information from EC-FP7 OpenAIRE currently gathers funding information from Wellcome Trust (UK) and FCT (Portugal) and extracts the links to projects from publication full text.

The EC and ERC are now able to monitor the scientific output within the FP7 programme. For the FP7 programme, OpenAIRE has identified nearly 100,000 publications, 52% of which are open access. For the ERC, OpenAIRE has identified 25,000 publications, of which 60% are open access.

from OpenAIRE via HighCharts (date: 6/2/2015)

Figure 2 - Ratio of closed vs open access publications in the FP7 funding stream

Applying the same methods to other national funders: OpenAIRE has worked closely with the EC to get detailed access to funding information, testing its results along the way. However each

funding agency has their own proprietary reporting and monitoring mechanisms. One approach OpenAIRE has adopted, when EC funding details cannot be found in the repository, is retrospective addition of funding information. OpenAIRE offers services to identify funding information. Text mining algorithms can identify funding information within abstracts, or the full text. The user then needs to claim the publication to the project. OpenAIRE is now experimenting with offering its services to other funders, and this paper presentation will also put forward the results of some pilots with other national funders, namely FCT (Fundação para a Ciência e Tecnologia), Portugal. The same services are being applied to the main national funder in Greece.

Conclusion

The EC has now asserted its OA policy in Horizon 2020, and OpenAIRE is able to support this by leveraging the valuable contents of repositories. Scholarly communication infrastructures could ease both monitoring and reporting processes and ensure that scientific results are automatically integrated without any additional burden to researchers and research managers. However there is work ahead to promote quality metadata and the inclusion of funding information in metadata records of institutional and disciplinary repositories to maintain their vital role in the open access movement.

References

[i] www.openaire.eu

[ii] http://ec.europa.eu/programmes/horizon2020

[iii] https://www.openaire.eu/oa-publications/h2020/open-access-in-horizon-2020

[iv] https://guidelines.openaire.eu/wiki/Main_Page