OpenAIRE: Advancing Open Science

Paolo Manghi[1], Michele Artini[1], Claudio Atzori[1], Miriam Baglioni[1], Alessia Bardi[1], Sandro La Bruzzo[1], Michele De Bonis[1], Harry Dimitropoulos[2], Ioannis Foufoulas[2], Katerina Iatropoulou[2], Natalia Manola[2], Stefania Martziou[2], Pedro Principe[3]

[1] Institute of Information Science and technologies, Italian National Research Council
[2] Athena Research Center in Information, Communication and Knowledge Technologies
[3] University of Minho, Portugal

Abstract OpenAIRE[1], the point of reference for Open Access in Europe, is now addressing the problem of enabling the Open Science paradigm. To this aim it will provide services to: (i) overcome the limits of today's scientific communication landscape, by allowing research communities and the relative e-infrastructures to fully publish, interlink, package and reuse their research artefacts (e.g. literature, data, and software) and their funding grants within the European and global ecosystem as supported/promoted by OpenAIRE, (ii) enable end-users (e.g. researchers, funder officers) to search and consult a rich and up-to-date knowledge graph of research results and (iii) enable scientific and educational information repositories and publishers to subscribe and be notified of changes in the OpenAIRE knowledge graph. These combined actions will bring long-term and immediate benefits to research communities, research organisations, repository managers, and funders by affecting the way research results are disseminated and reused. On the one hand, publishing the interlinked and packaged research literature, data and software via OpenAIRE drives research communities to an Open Science transition in a consistent and interoperable fashion. On the other hand, the resulting infrastructure concretely enables the construction of Open Science oriented services, supporting practices such as machine-assisted research reproducibility and evaluation.

1. Introduction

Open Science is frequently defined as an umbrella term that involves various movements aiming to remove the barriers for sharing any kind of output, resources, methods or tools, at any stage of the research process 9. This is intended as a means for accelerating research by enhancing transparency and collaboration, and fostering innovation and reproducibility. Scientists and organizations see Open Science as a way to speed up, improve quality, and more effectively reward research activities, while funders and ministries see it as a means to optimize cost of science and leverage innovation. Open Science is an emerging vision, a way of thinking, whose challenges always gaze beyond its actual achievements. Today, the effective implementation of Open Science calls for a scientific communication ecosystem capable of enabling Open Science publishing principles. The ecosystem should allow research communities to share (for discovery and transparent evaluation) and re-use (for reproducibility 213) their scientific results by publishing all intermediary and final research artefacts, beyond

---

[1] OpenAIRE www.openaire.eu

scientific literature. Artefacts can be research data, software and research methods (e.g. workflows, protocols, algorithms, etc.), which should be deposited in repositories for scientific communication (e.g. institutional repositories, data archives, software repositories, CRIS systems), and should be published together with the semantic links between them. To complete the picture, such ecosystem should support publishing of packages of artefacts (e.g. research objects 214, enhanced publications 515, RMap 16) to allow discovery, evaluation, and reproducibility of science (e.g. workflows or experiments with input datasets).

Today's scientific communication landscape is far from supporting this vision, mainly due to its inability to:

1. *Support publishing of all kinds of research artefacts.* For example, research methods publishing workflows are generally not best practice, i.e. no research method repositories, no persistent identifiers for methods, no citation practices and, therefore, no scientific reward;

2. *Keep a complete and up-to-date record of research artefacts relationships.* For example, publication, data, software repositories and publishers do not keep bi-lateral links between each other's artefacts, and the links they keep are not in-sync with the updates of the artefacts (e.g. links to new versions of the data, obsolete links);

3. *Find agreements on how to share and publish packages of artefacts.* Solutions exists (e.g. research objects, enhanced publications, RMap) but are specific to rather small communities of scientists, implying that research packages, as well as research methods, are not regarded as first-class citizens in the scientific communication domain.

De facto, today's scientific communication ecosystem lacks tools and practices for engaging research communities at adopting the aforementioned novel Open Science publishing principles, even when researchers are already in the position of publishing interlinked artefacts and/or packages.

OpenAIRE fosters transparent evaluation of results and facilitates reproducibility of science for research communities by enabling a scientific communication ecosystem where artefacts, packages of artefacts, and links between them can be exchanged across communities and across content providers. To this aim, OpenAIRE, via the OpenAIRE-Connect project, introduces and implements the concept of Open Science as a Service (OSaaS) on top of the existing OpenAIRE infrastructure 1, by delivering services in support of Open Science. Following the NIST definition of *aaS* service models[2], the service model "Open Science as a Service" provides the consumers with the capability of accessing tools that implement Open Science principles, transparently with respect to the underlying technical infrastructure. Tools are accessible through either a thin client interface, such as a web browser, or an application program interface (API).

OpenAIRE-Connect[3] will realize and operate two services for Open Science. The first, Research Community Dashboard, will serve research communities to publish research artefacts, packages, and links, and to monitor their research impact. The second, Catch-All Notification

---

Broker Service, will engage and mobilize content providers, and serve them with services enabling notification-based exchange of research artefacts, to leverage their transition towards the Open Science paradigm. Both services will be served on-demand according to the OSaaS paradigm, hence be re-usable by different disciplines and providers, each with different practices and maturity levels, so as to favor a shift towards a uniform cross-community and cross-content provider scientific communication ecosystem.

The adoption of these services, eased by the OSaaS approach, aims at incepting Open Science publishing within the existing scholarly communication landscape. By introducing an OSaaS approach, OpenAIRE-Connect will deliver on-demand Open Science publishing services to research communities and content providers, aligning practices and mechanisms that address transparent evaluation and reproducibility (see Figure 1). By complementing the technological efforts with networking activities that will strengthen the emerging Open Science social environment, OpenAIRE-Connect will facilitate a cultural and technological shift towards common Open Science publishing practices.

To achieve its objectives, OpenAIRE-Connect involves key stakeholders of scientific communication: a pool of forward-looking research communities, today publishing or in the need of publishing research data and methods, international representatives of Open Access publishers (Jisc 3), publication repositories (COAR 9), and data archives (ICSU World Data Systems/WDS 11), willing to support and benefit from such a change.
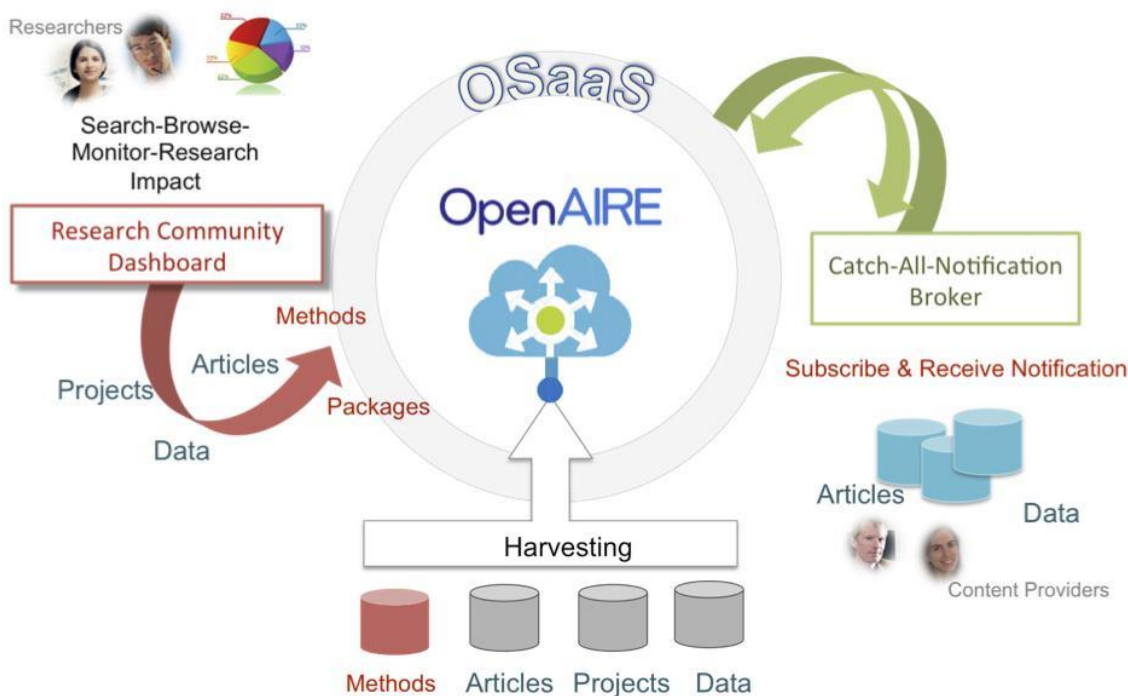

Figure 1. Research Community Dashboard and Catch-All Notification Broker Service

2. OpenAIRE services in support of Open Science

OpenAIRE-Connect extends the technological services today offered by the OpenAIRE infrastructure in order to foster the adoption of Open Science publishing practices and facilitate the emergence of shared solutions. Specifically, OpenAIRE-Connect introduces two classes of new services:

1. *Research community services* offering support for a uniform transition of research communities towards Open Science publishing via the Research Community Dashboard;
2. *Content provider services* leveraging the transition of content providers towards Open Science publishing via the Catch-All Notification Broker Service.

Continuing and building on OpenAIRE's openness and sharing of content, services and practices, OpenAIRE-Connect will develop a *uniform, common strategy* for approaching and engaging with research communities (especially Research Infrastructures targeting ESFRI 8 programmes), which will be a major outcome of the project. This strategy will be used in due course by the wider OpenAIRE constituency, i.e. the National Open Access Desks (NOADs), in its strategic synergies within the emerging European Open Science Cloud (EOSC 12) ecosystem to increase awareness on the Open Science topics and to promote the adoption and uptake of the new services.

In synergy with EOSC, towards the widely discussed need of the gluing social aspect (support and human infrastructure), OpenAIRE-Connect aims to design and deliver *a targeted support and training programme for research communities* and relevant stakeholders. This will inform them about the benefits and use of the OpenAIRE-Connect services, to pass on best practices and to lower the barriers of participation in the Open Science ecosystem and, particularly, in the OpenAIRE infrastructure.

2.1. Research Community Dashboard

OpenAIRE-Connect will support the evolution of today's fragmented scientific communication landscape by providing researchers of specific communities with services giving access to facilities for collaboratively maintain an up-to-date knowledge graph of their interlinked or packaged research artefacts, e.g. literature, data, software and methods. This research community graph will be built as an extension of the broader knowledge graph today populated by the OpenAIRE infrastructure by adding software and other research artefacts (all the artefacts that are different from literature, dataset, and software). It will therefore integrate and inherit links to funders, projects, literature and datasets as inferred from article full-texts (today more than 4 millions) or harvested by OpenAIRE from content providers (today more than 1000). These facilities will be provided by a new infrastructural service, the OpenAIRE Research Community Dashboard, that each community will be able to request and configure according to its specific needs. Each dashboard will serve two types of community users (researchers and research community operators, who have an "administrative role" for the community and can configure the dashboard) with a suite of value added functionalities:

- *Connect and Link*: researchers authoritatively provide and curate links between artefacts related to their scientific community, a process moderated by the research community operators.

- *Deposit*: researchers who have not a repository of reference can deposit in OpenAIRE's Zenodo 67 files and metadata relative to their research literature, data, software, methods, and packages, and obtain a DOI.
- *Enrich Content* (configurable inference): research community operators authoritatively tune the configuration of OpenAIRE text mining algorithms with community specific rules to identify artefacts relevant for the research community;
- *Learn your Impact*: the research community can reliably monitor and report the research impact of their scientific production w.r.t. several European (and beyond) funders, visualize trends, classifications resulting from OpenAIRE's analytics services; the knowledge graph inclusive of all community artefacts, with links between them and relative attribution of work, enables transparent evaluation of science.
- *Discovery and reuse*: researchers can restrict their search, browse, and navigation focus to the subpart of the OpenAIRE information space associated to the community; discovering and accessing packages of artefacts fosters reproducibility of science.

OpenAIRE-Connect will also develop, in collaboration with other international initiatives, interoperability guidelines to enable the exchange of research artefacts and packages and will offer APIs enabling third-party services to bulk-feed research artefacts into the OpenAIRE knowledge graph according to the established guidelines.

OpenAIRE-Connect involves its onset researchers from a wide range of communities looking into pragmatic solutions to research data and methods publishing in Open Science settings. With different levels of maturity, touching upon interdisciplinary, they will deliver end-user requirements for the realization of the *Dashboard* service, and engage in pilots to test, assess, and adopt the services:

- *Earth and Environmental Sciences (UniHB)*: the ATLAS[4] community relies on thematic data archives (Pangaea) whose datasets are packages of datasets and related literature, aiming to link to different scientific domains.
- *Neuroinformatics (CNRS):* the France Life Imaging national infrastructure[5] produces data images, links them with methods (software and services), and produces packages.
- *Fisheries and aquaculture management (CNR, IDR):* the BlueBridge[6] and MARBEC[7] infrastructures are moving towards collaborative editing of "dynamic publications", looking for Open Science solutions.
- *Humanities and Cultural Heritage (PIN):* PARTHENOS[8], a cluster of research infrastructures from Linguistics, Humanities, Cultural Heritage, History, Archaeology, with different types of data to interconnect.
- *Environment & Economics (ICRE8):* the national/EU node of the United Nations Sustainable Development Solutions Network[9] sets out to build an infrastructure to gather

---

[4] ATLAS: https://www.eu-atlas.org/

[5] France Life Imaging: https://www.francelifeimaging.fr/

[6] BlueBridge: http://www.bluebridge-vres.eu/

[7] MARBEC: http://www.umr-marbec.fr/en/?lang=en

[8] PARTHENOS: http://www.parthenos-project.eu/

[9] United Nations Sustainable Development Solutions Network: http://unsdsn.org/

all publications and data available in repositories and in Public Sector Information portals, and link them to European and national funding.

2.2. Catch-All Notification Broker Service

Research artefacts repositories, a.k.a. content providers, (e.g. institutional and thematic repositories, aggregators, data archives) are key in serving research communities towards their Open Science goals. To ease the adoption of Open Science principles among researchers it is important to lower the barriers to Open Science publishing.

The "publish/deposit once" practice is a decisive means in the overall Open Science roadmap, and will only be achieved when content providers seamlessly connect to the wider open scientific communication ecosystem. This will allow them to pro-actively exchange information about research artefacts and links of value to all interested communities or stakeholders, without the researchers having to worry *where*, *when*, and *how* to publish in order to fulfill the numerous mandates.

As part of the OSaaS portfolio, OpenAIRE-Connect will develop and deploy a Catch-All Notification Broker Service that connects all types of content providers (institutional repositories, publishers, data repositories, and CRIS systems). The Catch-All Notification Broker will extend OpenAIRE's notification brokering service 4, which serves literature repositories, and will broaden the content provider base with the ones that serve specific research communities.

Thanks to its functionality, providers can be notified of metadata records relative to artefacts that are "of interest to them" (i.e. metadata records that should be in the content provider's data base), or "linked to them" (i.e. a scholarly link exists between one of the provider's artefact and the identified artefact).

Notifications are sent only to subscribed providers, following a subscription and notification pattern, and can be delivered by mail, OAI-PMH interfaces, or, currently under investigation, via push APIs (e.g. SWORD protocol), FTP and ResourceSync.

This will effectively allow content providers to complete or enrich their collection of artefacts with up-to-date information from the wider OpenAIRE ecosystem, and research communities or infrastructures to have a direct communication line with content providers via OpenAIRE.

The idea behind the service is to disseminate and advocate the principle that scholarly communication data sources are not a passive component of the scholarly communication ecosystem, but rather active and interactive part of it. They should not consider themselves as thematic silos of products, but rather as hubs of products semantically interlinked with any kinds of research artefacts and, more broadly, up-to-date with the evolving research ecosystem.

OpenAIRE-Connect brings on board leading representatives of institutional repositories (e.g. COAR), data repositories (UniHB/Pangaea, UniHB/WDS), publishers (via Jisc) that are already moving towards Open Science-oriented publishing and are committed to provide requirements and engage in the experimentation of the brokering services. Beyond the ones on board to the project, i.e. Pangaea (UniHB) and Zenodo (CERN), a number of content providers have already indicated their interest in the Catch-All Notification Broker Service and are ready to engage in a number of pilots for testing and adopting the service: the German GESIS Datorium repository for datasets and scripts in the Social Sciences, the FCT-FCCN network of Portuguese

institutional repositories, Australian ANDS data archive, and Open Access publishers, such as eLife, Frontiers, EuropePMC.

In addition, OpenAIRE-Connect will implement a pilot for exchanging subscriptions and notifications between the Catch-All Broker Service and the Jisc UK Notification Router, in order to serve each other's "customers" with a wider range of subscriptions and notifications (see Figure 2). This cooperation goes in the direction of a scientific communication ecosystem where brokers conforming to common requirements for the exchange of subscriptions and notifications can interoperate to collaboratively (by granting or delegating subscriptions to the network) channel the information they collect from producers of events (OpenAIRE for the Catch-All Broker Service and publishers for the Jisc Notification Router) to interested consumers (e.g. any content provider in OpenAIRE and UK repositories for Jisc Notification Router).
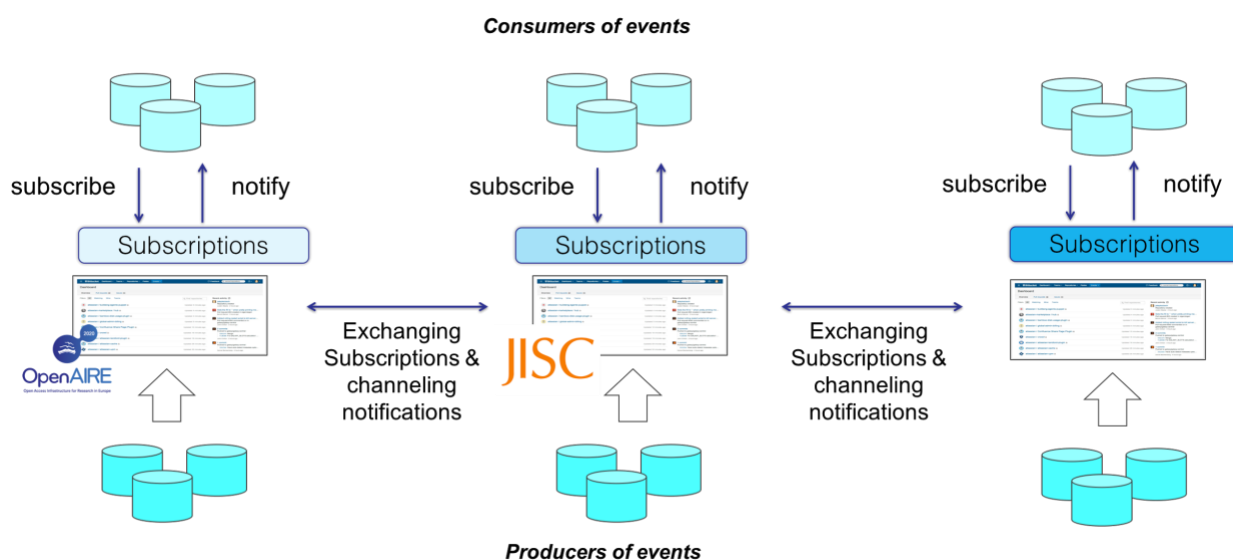

Figure 2. Interoperability between the Broker services

3. Conclusions

The effective implementation of Open Science calls for a scientific communication ecosystem capable of enabling the "Open Science publishing principles" of transparency and reproducibility. Such ecosystem should provide tools, policies, and trust needed by scientists for sharing and interlinking (for "discovery" and "transparent evaluation") and re-using (for "reproducibility") all research artefacts produced during the scientific process, e.g. literature, research data, methods, software, workflows, protocols.

OpenAIRE fosters Open Science by advocating its publishing principles across Europe and research communities and by offering technical services in support of Open Access monitoring, research impact monitoring, and Open Science publishing. Its aim is to provide Research Infrastructures (RIs) with the services required to bridge the research life-cycle they support (where scientists produce research artefacts) with the scholarly communication infrastructure (where scientists publish research artefacts) in such a way science is reusable, reproducible,

and transparently assessable. OpenAIRE is fostering the establishment of reliable, trusted, and long lasting RIs by compensating the lack of Open Science publishing solutions and by providing the support required by RIs to upgrade existing solutions to meet Open Science publishing needs (e.g. technical guidelines, best practices, Open Access mandates). To this aim, OpenAIRE is working closely with existing RIs to extend its portfolio by implementing the concept of "Open Science as a Service" (OSaaS) and offer two new services: the Research Community Dashboard and the Catch-All Notification Broker Service.

The first beta release of the services is planned on March 2018. A set of testing sessions will be conducted by five research communities (for the Research Community Dashboard) and a number of content providers (for the Catch-All Notification Broker Service) before the first production public release, expected in June 2019.

Acknowledgments

References

1. Manghi, P., Bolikowski, L., Manold, N., Schirrwagen, J., & Smith, T. (2012). OpenairePlus: the european scholarly communication data infrastructure. *D-Lib Magazine*, 18(9), 1. http://doi.org/10.1045/september2012-manghi
2. Bechhofer, Sean, Iain Buchan, David De Roure, Paolo Missier, John Ainsworth, Jiten Bhagat, Philip Couch et al. "Why linked data is not enough for scientists." Future Generation Computer Systems 29, no. 2 (2013): 599-611. https://doi.org/10.1016/j.future.2011.08.004
3. Jisc Publication Router http://www.jisc.ac.uk
4. Artini, Michele, Claudio Atzori, Alessia Bardi, Sandro La Bruzzo, Paolo Manghi, and Andrea Mannocci. "The OpenAIRE literature broker service for institutional repositories." *D-Lib Magazine* 21, no. 11/12 (2015). http://doi.org/10.1045/november2015-artini
5. Hoogerwerf, Maarten, Mathias Lösch, Jochen Schirrwagen, Sarah Callaghan, Paolo Manghi, Katerina Iatropoulou, Dimitra Keramida, and Najla Rettberg. "Linking data and publications: towards a cross-disciplinary approach." *International Journal of Digital Curation* 8, no. 1 (2013). http://doi.org/10.2218_ijdc.v8i1.257
6. *Zenodo.* https://zenodo.org
7. Potter, M., & Nielsen, L. H. (2015, September 1). Zenodo Information Architecture and Usability. http://doi.org/10.5281/zenodo.31870
8. ESFRI https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri
9. COAR repository https://www.coar-repositories.org
10. FOSTER https://www.fosteropenscience.eu/node/1420
11. WDS https://www.icsu-wds.org
12. EOSC https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

13. Goodman, S. N., Fanelli, D., & Ioannidis, J. P. (2016). What does research reproducibility mean?. Science translational medicine, 8(341). http://doi.org/10.1126/scitranslmed.aaf5027

14. Research Object http://www.researchobject.org

15. Bardi, A., & Manghi, P. (2014). Enhanced Publications: Data Models and Information Systems. *LIBER Quarterly*, *23*(4), 240–273. http://doi.org/10.18352/lq.8445

16. RMap. http://rmap-project.info