

The OpenAIRE Research Community Dashboard: on Blending Scientific Workflows and Scientific Publishing

Miriam Baglioni¹, Alessia Bardi¹, Argiro Kokogiannaki², Paolo Manghi¹,
Katerina Iatropoulou²

Pedro Principe³, André Vieira³, Lars Holm Nielsen⁴, Harry Dimitropoulos²,
Ioannis Foufoulas², Natalia Manola², Claudio Atzori¹, Sandro La Bruzzo¹,
Emma Lazzeri¹, Michele Artini¹, Michele De Bonis¹, Andrea Dell’Amico¹

¹ Consiglio Nazionale delle Ricerche,
Istituto di Scienza e Tecnologie dell’Informazione “A. Faedo”, Pisa, Italy
`name.surname@isti.cnr.it`

² Department of Informatics and Telecommunications
National and Kapodistrian University of Athens
{argirok, kiatrop, natalia, harryd, i.foufoulas}@di.uoa.gr

³ Universidade do Minho, Pisa, Italy
pedroprincipe@s dum.uminho.pt, andrevieira@s dum.uminho.pt

⁴ CERN IT Department, Geneva, Switzerland
lars.holm.nielsen@cern.ch

Abstract. Despite the hype, the effective implementation of Open Science is hindered by several cultural and technical barriers. Researchers embraced digital science, use “digital laboratories” (e.g. research infrastructures, thematic services) to conduct their research and publish research data, but practices and tools are still far from achieving the expectations of transparency and reproducibility of Open Science. The places where science is performed and the places where science is published are still regarded as different realms. Publishing is still a post-experimental, tedious, manual process, too often limited to articles, in some contexts semantically linked to datasets, rarely to software, generally disregarding digital representations of experiments.

In this work we present the OpenAIRE Research Community Dashboard (RCD), designed to overcome some of these barriers for a given research community, minimizing the technical efforts and without renouncing any of the community services or practices. The RCD flanks digital laboratories of research communities with scholarly communication tools for discovering, publishing, and monitoring interlinked scientific products such as literature, datasets, and software. The benefits of the RCD are show-cased by means of two real-case scenarios: the European Marine Science community and the European Plate Observing System (EPOS) research infrastructure.

Keywords: Open Science · scholarly communication · research infrastructures · research communities

Note This is the pre-print of an article published at TPDL conference 2019. Please cite as: *Baglioni M. et al. (2019) The OpenAIRE Research Community Dashboard: On Blending Scientific Workflows and Scientific Publishing. In: Doucet A., Isaac A., Golub K., Aalberg T., Jatout A. (eds) Digital Libraries for Open Knowledge. TPDL 2019. Lecture Notes in Computer Science, vol 11799. Springer, Cham*

1 Introduction

Research communities are increasingly operating under the hat of *Research Infrastructures* (RIs), intended as initiatives of governance where scientists and organizations can define common policies, recommend best practices, and deliver and maintain digital services to leverage better scientific workflows and foster innovation. Within RIs researchers can assemble their *digital laboratory* by selecting the services, tools and resources they need, and perform digital science according to common scientific workflows and sharing of scientific results. Digital laboratories are the place where researchers perform their digital experiments, share and find scientific products, track their lineage and provenance and, where possible, enable their fully-fledged re-use. In fact, based on the degree of maturity of a digital laboratory, researchers may find the R^* conditions to repeat (“same research activity, same lab”), replicate (“same research activity, different lab”), reproduce (“same research activity, different input parameters”), or re-use (“using a product of a research activity into another research activity”) the digital experiment, thereby maximizing transparency and exploitation of scientific findings [6].

Still, although digital science and digital laboratories facilitate the way to Open Science, a number of cultural and technical barriers are to be overcome before this vision will be effectively achieved. One of the grand challenges is the clear cultural and technical separation between digital laboratories and the scholarly communication ecosystem, intended as the environment used by science stakeholders (e.g. researchers, organizations, funders, academic societies, publishers) to publish, review, find, evaluate, and monitor science.

The mission of the OpenAIRE initiative¹, one of the pillars of the European Open Science Cloud (EOSC)², is to address this challenge at both cultural and technical level, by providing training, dissemination, and technical services to incept Open Science publishing practices into the research life-cycle. In this work, we shall present the OpenAIRE Research Community Dashboard (RCD), a service conceived to provide research communities with transparent bridges between their digital laboratory and the scholarly communication ecosystem. The RCD provides tools for scientists to discover, interlink and manually or automatically publish all kinds of research products and for officers to monitor the Open Science, Open Access, scientific trends and statistics of the community. This complements the mission of research infrastructure services, minimizing the

¹ OpenAIRE: www.openaire.eu

² EOSC: <https://www.eosc-portal.eu>

cost of publishing, while respecting community practices and relative scientific workflows. Two use-cases will be presented to illustrate the RCD functionalities, relative to the European Plate Observing System (EPOS) infrastructure³ and the European Marine research community⁴.

Outline Section 2 describes the barriers to Open Science publishing identified in the gap between digital laboratories and scholarly services. Section 3 describes the dissemination, training, and technical solutions to such challenges proposed by the OpenAIRE infrastructure, while Section 4 describes the Research Community Dashboard and its functionalities. Finally, Section 5 reports on the real-case scenarios of RCD adoption for the community of European Marine Science and the EPOS Research Infrastructure. Related work is illustrated in Section 6, while Section 7 concludes the paper and discusses future work.

2 Digital laboratories and scholarly communication ecosystem: the gap

According to Open Science (OS) principles [7, 8], all scientific products generated by research activities (e.g. scientific literature, research data, software, experiments) should be as open access as possible (“as closed as necessary”), made available as soon as possible, and “under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods”.⁵ To implement this vision, scientists should follow OS publishing workflows while being technically and properly supported by digital laboratories, whose services and tools should not only allow for scientific processes to be carried out, but also for the resulting digital products to be published via scholarly/scientific communication services. The European Open Science Cloud is pro-actively promoting this vision and supporting communities and research infrastructures at implementing it, but still the overall level of technical and cultural maturity across RIs is rarely sufficient to achieve the R^* conditions. Surveying RI experiences in collaboration with OpenAIRE (see Section 5), we can observe common lacks:

- *Article-centric publishing*: many research communities still show poor interest in publishing research products beyond the scientific article; when other products are published, research data is the focus, with links to articles, while software publishing (preservation, DOIs, etc) is still largely overlooked, and digital experiment publishing (e.g. methods, workflows, research objects) is an exception;
- *Article-centric scientific reward practices*: poor or absent support for scientific reward systems that take into account all kinds of research products;
- *Fragmentation of research products*: products are often published into community-independent data sources, where they lose their “community flavour”;

³ European Plate Observing System (EPOS): <https://www.epos-ip.org>

⁴ EuroMarine: <https://www.euromarinenetwork.eu>

⁵ FOSTER definition of Open Science: <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>

- identifying the collection of products of one community across such sources is in many cases not straightforward or requires manual/technical high costs;
- *Manual publishing cost*: publishing is mostly a manual action, there is poor or absent support for on-demand, automated publishing of research objects produced via services of RIs;
 - *Static publishing*: publishing occurs at the end of the research activity, as an act of “freezing” the products by deposition onto scholarly communication data sources; any event taking place after publishing to the products (e.g. citation, new version, usage in another experiment) is not dynamically reflected/materialized in the data sources with proper semantic links.

In the following we shall present the solutions undertaken by the OpenAIRE infrastructure in trying to blend these two worlds.

3 The OpenAIRE infrastructure

For a sustainable and smooth implementation of Open Science publishing, the places where research is performed, i.e. the digital laboratories, and the place where research is published, i.e. the scholarly communication ecosystem, should be bridged. The mission of the OpenAIRE initiative, one of the pillars of the European Open Science Cloud (EOSC), is to address this challenge at both cultural and technical level, by providing training, dissemination, and technical services to incept Open Science publishing practices into the research life-cycle. OpenAIRE is a Legal Entity representing a partnership of more than 50 institutions, working to shape and implement effective Open Access and Open Science policies in support of FAIR research, reproducible science, transparent assessment and omni-comprehensive evaluation.

OpenAIRE networking services OpenAIRE supports the implementation and alignment of Open Science policies at the international level by developing and promoting the adoption of global open standards and interoperability guidelines⁶ to realize a sustainable, participatory, trusted, scholarly communication ecosystem, open to all relevant stakeholders (e.g. research communities and organizations, funders, project coordinators) and capable of engaging society and foster innovation. The network of 34 National Open Access Desks (NOADs)⁷, present in every European country (and beyond) increases the awareness at the local and national level, reaching out to researchers, project coordinators and policy makers with training and support activities like workshops and webinars.

OpenAIRE technical services The monitoring of Open Science publishing trends is supported by a big data service infrastructure. The core of such technical infrastructure consists of metadata aggregation services that collect metadata

⁶ OpenAIRE Guidelines for Content Providers: <https://guidelines.openaire.eu>

⁷ OpenAIRE National Open Access Desks: <https://www.openaire.eu/contact-noads>

records relative to digital research products (literature, dataset, software, and others) from more than 13,000 scholarly data sources world-wide, for a current total of more than 30 million metadata records. Aggregated metadata records are represented in the form of a directed labelled graph where each record is a node (with properties) and each semantic link is a labelled edge, where the label expresses the semantics of the relationship (e.g. the dataset is *supplement to* a journal article, the publication is *funded by* a project). The graph is also processed by algorithms to (i) find and merge metadata records that describe the same object [2], (ii) apply inference techniques on the metadata records and full-texts of Open Access publications to add new properties and new semantic relationships [11].

The resulting graph is called the OpenAIRE Research Graph, on top of which added-value services are built to serve different scholarly communication stakeholders. The Research Graph data model is described in details in [12] and

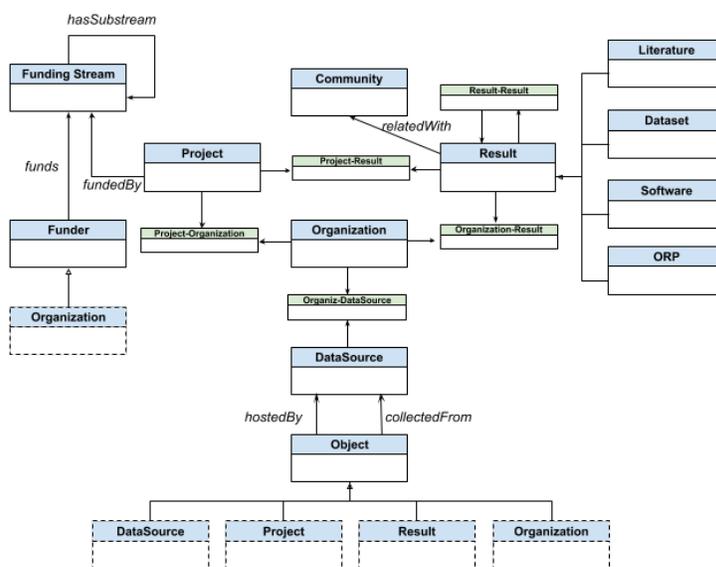


Fig. 1. OpenAIRE data model

depicted in figure 1. The main entities of the model are:

Results: outcome of research activities, which in OpenAIRE are of four classes *Datasets*, *Software*, *Literature* and *Other Research Products* (ORP). They are related to *Projects*, *Organizations*, *Communities*, and other *Results*.

Funders: *Organizations* (e.g. European Commission, Wellcome Trust, FCT Portugal, Australian Research Council) responsible for a list of *Funding Streams* (e.g. FP7 and H2020 for the EC).

Funding Streams identify the strands of investments managed by a *Funder* and can be nested to form a tree of sub-funding streams (e.g. FP7-IDEAS, FP7-HEALTH).

Projects are research projects under a *Funding Stream* of a *Funder*. *Projects* are associated to the *Results* produced under their activities and grants.

Data sources: web sources from which OpenAIRE collects the metadata (and the full-texts, if possible) of the objects in the graph. Each object is associated to the *Data source* from which the metadata was collected and to the one where the object is hosted; e.g. articles whose metadata is collected from DOAJ⁸ have PDFs hosted by specific journals.

Communities: include *research communities* and *research infrastructures*. Research communities are intended as communities of practice in a research field, willing to share and discover scientific results among the community itself and beyond. Research infrastructures are intended as service providers of research communities that are willing to monitor their impact in terms of results produced thanks to their existence.

The Research Graph is openly available under CC-BY license⁹ to programmatic clients via the DEVELOP API and to humans via the EXPLORE portal (<https://explore.openaire.eu>). Funders, officers of research initiatives, project coordinators and policy makers can use the functionality of the MONITOR portal (<https://monitor.openaire.eu>) to access statistics about the scientific production, monitor the uptake of Open Science publishing practices and use tools that support official reporting. The PROVIDE dashboard (<https://provide.openaire.eu>) target data source managers (e.g. repository managers) by offering tools to register data sources in OpenAIRE, validate their metadata records against the OpenAIRE guidelines, subscribe to other OpenAIRE added-value services like the OpenAIRE Usage Statistics¹⁰ and to the Catch-all Broker[1]. Finally, via the CONNECT portal (<https://connect.openaire.eu>) research communities and infrastructures can request the Research Community Dashboard, which offers a set of Open Science tools described in details in Section 4.

4 The OpenAIRE Research Community Dashboard

The RCD has been realized in the context of the EC H2020 OpenAIRE-Connect project, in collaboration with a number of research communities, which provided

⁸ DOAJ (Directory of Open Access Journals)

⁹ Some data sources aggregated by OpenAIRE do not allow to re-distribute metadata in CC0; the availability of another version of the graph including only information that can be redistributed in CC0 is in OpenAIRE's plans.

¹⁰ Guide about the OpenAIRE Usage Statistics Service: <https://www.openaire.eu/guides-usage-statistics>

requirements and tested the service for improvement and usability. It has been designed to address the challenges identified in Section 2 by means of the following strategies: (i) building a single entry point, called Community Gateway, where scientists can find all research products and links between them; (ii) provide community researchers with customizable Open Science publishing tools; (iii) enabling the customization of the criteria to be applied by the OpenAIRE mining and tagging algorithms to associate products in the OpenAIRE Research Graph to the community; (iv) providing APIs for research infrastructure services, part of digital laboratories, to automatically publish research products (e.g. datasets, software, digital experiments) in Zenodo, the OpenAIRE “catch-all” repository, and associate them to the community thanks to the concept of Zenodo community, a “container” of deposited products of the community¹¹. The RCD can serve two types of customers: research infrastructures in need of tools for monitoring and reporting the outcome of the science they support and research communities willing to find a scientific gateway for their Open Science publishing needs.

For scientists The gateway offers a discovery portal where users can search within the research products relative to their community. Moreover, then find publishing-oriented tools to: (i) manually deposit and get a persistent identifier (DOI) for research products of any kind (e.g. datasets, software, literature, methods, workflows, research packages) via Zenodo; (ii) interlink existing research products, e.g. link a dataset to the software that produced it; (iii) claim (i.e. assert) that a given research product, currently available in the OpenAIRE Research Graph, but not in the community gateway, is relevant for the community. Via such tools, researchers populate and access an open, participatory scholarly communication graph of interlinked objects through which they can share any kind of products in their community, maximise re-use and reproducibility of science, and outreach the scholarly communication at large.

For RCD managers An RCD is managed by a set of *RCD managers*, which can configure the mining and tagging algorithms used by the OpenAIRE infrastructure to assign scientific products in the OpenAIRE Research Graph to the community. OpenAIRE algorithms analyze the research graph to identify all products that are relevant for the community (e.g. products with relevant subjects, deposited in thematic or community specific repositories and archives, linked to domain-specific projects). In particular, RCD managers can perform the following actions (in any order):

- Select projects (among those available in OpenAIRE) that are related to the community: all research products linked to those projects will be automatically discoverable via the gateway.

¹¹ Zenodo communities: <https://zenodo.org/communities/>

- Select data sources (among those available in OpenAIRE) that are relevant for the community: all research products collected from such sources will automatically be discoverable via the gateway.
- Select Zenodo communities related to the community: research products in the given Zenodo communities will be discoverable via the gateway.
- Specify keyword/subject terms related to the community: all the research products whose metadata matches such terms will be discoverable via the gateway.
- Define full-text mining rules: all publications whose full-text matches the rules will automatically be discoverable via the gateway; e.g., some research infrastructures mandate specific acknowledgment statements when research was supported by infrastructure services.

In addition, RCD managers can also:

- Manage end-users’ claims: researchers can via UIs assert that a given product is relevant to the community. The RCD manager can confirm or reject those claims, acting as a “community moderator”.
- Configure statistics and charts: OpenAIRE makes available a number of stats for the community (e.g. percentage of Open Access publications and data, number of publications linked to data and to software, growth of Open Access publishing through the years). The RCD managers can decide which stats are public and which are private, for internal monitoring.

For RI Services Cultural barriers represent one of the major obstacles to the implementation of Open Science publishing. Digital products may well be generated by scientists via RI services, but still remain unpublished due to “lack of scientific reward” and “manual publishing cost”. In such cases, the OpenAIRE RCD can be used to support the RI at overcoming its current limits with minimal efforts and without renouncing any of its services or current practices. RI services can interact with the RCD to deposit scientific products through the Zenodo APIs, hence publish them as scholarly communication first-citizens, with attribution/citation metadata and a DOI. Community-specific metadata can be included in the deposition, so that researchers of the same domain can better interpret the scientific product, reproduce the activity or re-use the published results.

5 Real-case scenarios

The RCD is a BETA service (<https://beta.connect.openaire.eu>) supporting community gateways for 6 research communities (European Marine Science, Neuroinformatics, Digital Humanities and Cultural Heritage, Fisheries and Aquaculture Management, Sustainable Development Solutions Network Greece, Agricultural and Food Sciences) and 7 research infrastructures and initiatives (CLARIN, EGI, DARIAH-EU, ELIXIR-GR, Instruct-ERIC, Research Data Alliance, EPOS-IT),

whose members actively contributed to its testing, both in terms of functionality and GUI usability.¹² The RCD operates on the BETA OpenAIRE Research Graph, which aggregates today (June 2019) +400Mi metadata objects, counting, once deduplicated, +90Mi literature records, 8Mi dataset records, 140k software records, and 350Mi links between them objects (<https://beta.explore.openaire.eu>).

In the following we report the experience with two use-cases of RCD adoption: an Open Science gateway for the European Marine Science community and a Open Science monitoring gateway for the EPOS research infrastructure that implements on-demand publishing functionality.

5.1 The European Marine Science RCD Gateway

In Europe, the marine science community uses public and commercial infrastructures and services to publish and share their research products, mainly literature and data. These include about 150 scientific journals, a network of 45 national oceanographic data centres (e.g. SeaDataNet), a European Marine Observation and Data portal (EMODnet), operational data access services (e.g. Copernicus), geographic and taxonomic registers (e.g. MarineRegions and WoRMS), and thematic data archives that work closely with the research community (e.g. PANGAEA, EMBL-EBI and ICES).¹³ Efforts to connect and link these resources are growing, but they remain poorly known or used by the research community. Additionally, there is an important lack of infrastructure to store and share research methods such as model codes, model outputs, scientific reports and outreach/training material.

The European Marine Science RCD (<https://beta.mes.openaire.eu>) offers to the community of the EuroMarine network (a consortium of 56 research and academic organisations) the set of Open Science-oriented functionalities they need. Thanks to the community gateway, researchers have one single entry point where they can find all the research products relevant to their research discipline. Today (June 2019), they have access to more than 120,000 research products among publications, datasets, software and other research products associated to the community with no effort but the configuration process. The European Marine Science gateway has been configured to include products from about 50 Zenodo communities, 600 projects funded by the European Commission and Fondation Tara Expéditions¹⁴, some of the most relevant data sources for the community including PANGAEA and DRYAD, and finally a set of generic keywords in order

¹² Some of gateways are still in private mode and cannot be accessed.

¹³ Relevant resources for the Marine Science community: SeaDataNet (<https://www.seadatanet.org>), EMODnet (<http://www.emodnet.eu/>), Copernicus (<https://www.copernicus.eu>), MarineRegions (<http://www.marineregions.org/>), World Register of Marine Species (WoRMS) (<http://www.marinespecies.org/>), PANGAEA. Data Publisher for Earth and Environmental Science (<https://www.pangaea.de/>), European Bioinformatics Institute (EMBL-EBI, <https://www.ebi.ac.uk/>), International Council for the Exploration of the Sea (ICES, <https://www.ices.dk>), DRYAD (<https://datadryad.org>)

¹⁴ Fondation Tara Expéditions: <https://oceans.taraexpeditions.org>

to minimize the number of false negatives (i.e. results that are supposed to be in the gateway but they are not).

The RCD manager also decided to publicly show all available statistics and charts, even though numbers and pictures should be considered provisional because the configuration of the OpenAIRE algorithms has not yet been finalised. Figure 2 shows some screenshots of the GUI for European Marine Science RCD managers.

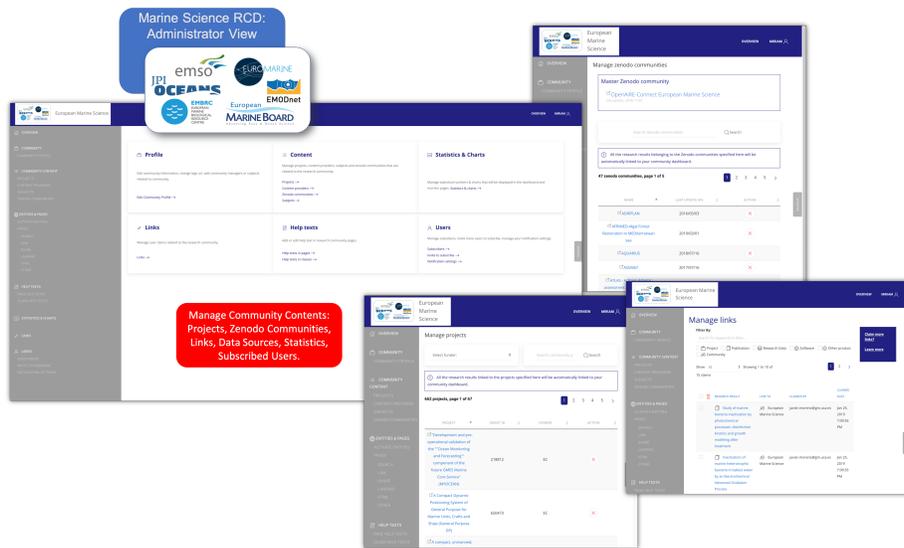


Fig. 2. RCD managers user interfaces

The European Marine Science RCD was initially created as a pilot in the context of the H2020 project OpenAIRE-Connect to show to researchers in the EuroMarine network the potential and the benefit of Open Science publishing practices. Thanks to the successful pilot, plans for the future include the deployment of additional RCDs for monitoring research infrastructures in the domain (e.g. EMBRC¹⁵ and EMSO-ERIC¹⁶) and advisory boards (e.g. ICES¹⁷).

5.2 The EPOS Research Infrastructure RCD Gateway

The European Plate Observing System (EPOS) is the pan-European distributed Research Infrastructure for solid Earth science to support a safe and sustainable

¹⁵ European Marine Biological Resource Centre (EMBRC): <http://www.embrc.eu/>

¹⁶ European Multidisciplinary Seafloor and water column Observatory (EMSO): <http://emso.eu>

¹⁷ The International Council for the Exploration of the Sea (ICES): <https://www.ices.dk>

society. Through the integration of National research infrastructures and data, EPOS will allow scientists to make a step change in developing new geo-hazards and geo-resources concepts and Earth science applications to help address key societal challenges. The ESAs Geohazards Exploitation Platform (GEP) is a system that offers as-a-service to EPOS scientists a number of scalable and parallel processing/analysis tools to be run as parallel jobs on a public/private cloud over user-identified satellite big datasets (e.g. from Copernicus). GEP tracks the history of execution of the algorithms and stores in a local database the output, in such a way past executions, i.e. GEP experiments, can be subsequently re-executed by GEP users.

According to Open Science publishing, the GEP should offer scientists the possibility to publish GEP experiments (i.e. algorithms, parameters, links to input data.) as well as the digital results. In a collaboration with EPOS, between the EOSC-Hub and OpenAIRE-Advance projects, OpenAIRE will offer an RCD for EPOS and provide support for GEP on-demand publishing [13] for the specific use-case of the CNR-IREA EPOSAR tool in GEP.

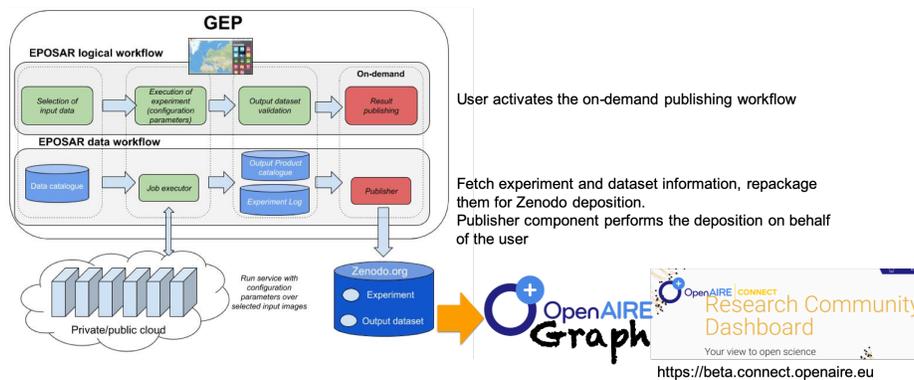


Fig. 3. EPOSAR workflow integrated with OpenAIRE RCD

Figure 3, shows how the current EPOSAR workflow implemented in GEP can be extended with an on-demand publishing process. The red boxes in figure 3 represent the steps that are missing to bridge the researchers' digital laboratory and the scholarly communication ecosystem. The EPOSAR GUI will be modified to prompt the researcher with the authorization to publish. GEP will be equipped with a publishing component capable of fetching experiment and dataset information from the local databases and deposit them as products (via APIs) in Zenodo.org under the EPOS community, on behalf of the EPOSAR authorized service. Such products will be given a DOI, described by citation/attribution metadata, reciprocally linked, equipped with semantic links to other products (e.g. the DOI of the articles describing the EPOSAR algorithms). As a consequence, scientists will publish all their scientific products if needed, and, as

a consequence of being deposited in Zenodo under the EPOS community, be discoverable/browsable through the OpenAIRE EPOS RCD. When extended to all GEP algorithms, the EPOS infrastructure will be able via the RCD to monitor/report all scientific products mature enough to be published. Moreover, experiments will reach out other scientists via the scholarly communication ecosystem, allowing them to access GEP and reproduce GEP experiments, as well as citing their authors when this is the case.

6 Related Work

The call for the implementation of Open Science principles generated a number of tools that facilitate their adoption.

SDSC ScienceGateways [14], VRE4EIC [10], and D4Science [4] are platforms for the deployment of Scientific Gateways or Virtual Research Environments (VREs), intended as web-based working environments where researchers of a community of practice can find services to perform their research activities and share intermediate and final results by guaranteeing ownership, provenance and attribution. [5] The VRE managers can select the tools and resources available in the platform that should be accessible in the VRE. Researchers use the VRE as their digital laboratory, as a single entry point to the resources relevant for their community, and as a dissemination means for their (ongoing) research activities. VREs typically focus on the OS principle of open collaboration and on the *R**s of the scientific process [3]. In the majority of the cases, VREs are decoupled from the scholarly communication infrastructure and must be flanked with tools like the Research Community Dashboard.

Another platform that supports the deployment of community-specific gateways for discovery and re-use purposes is SciCrunch¹⁸. SciCrunch was “designed to help communities of researchers create their own portals to provide access to resources, databases and tools of relevance to their research areas”. [9] It offers a high degree of customization of community portals in terms of content, look and feel and services. Communities can select relevant data sources from a registry of 250+ data repositories and 13,000 research resources in the biomedical domain. New sources can also be added and those become available for selection to all other communities using the platform. The look and feel and the search functionality of the community can be configured to exploit community-specific ontologies.

The approach of SciCrunch is very similar to the one of the OpenAIRE RCD. Both are designed to provide a single entry point to all research resources the researchers of a given community many need. This information is selected by community managers from a set of available sources that may change over time, enabling the possibility to include new community-specific sources. The main differences among SciCrunch and the RCD are to be found in the underlying content they are built upon and in the functionality offered to their users. SciCrunch works on biomedical data and can offer a higher degree of customization

¹⁸ SciCrunch: <https://scicrunch.org/>

of the gateways, since it serves different communities related to the same high-level discipline. The OpenAIRE RCD, instead, is built on top of the OpenAIRE research graph and can potentially be used to deploy gateways for research communities of any disciplines. Since the OpenAIRE research graph is not specific for any disciplines, the RCD offers more advanced configuration options in terms of content: the selection of the relevant data sources is only one of the options that the RCD managers can configure to include research products in their gateways. In addition, the RCD offers Open Science publishing tools that allows researchers to “interact” with the scholarly communication ecosystem from within the gateway in such a way the data sources responsible for the products can get notified about the new links they add (via the OpenAIRE Broker service).

7 Conclusions and future work

In this paper we have presented the OpenAIRE Research Community Dashboard (RCD) as a toolkit to support the shift of research communities towards Open Science publishing. The RCD has already passed two functional and usability testing phases to which members of different research communities and infrastructures have participated and is now ready for production. Future developments include (i) a general re-styling of the GUIs to improve the user experience; (ii) the possibility for RCD managers to configure gateway pages to provide community-specific recommendations for the adoption of OS principles (e.g. preferred repositories, publishing practices); (iii) the addition of the *tagging propagation mode*, allowing RCD managers to opt for research products to be associated to the community if they are semantically linked to products of the community; e.g. if an article is associated to the community and linked to a dataset with a *supplementedBy* relationship, then also the dataset can be associated to the community; and (iv) the integration with the Search and Link Wizard of ORCID, to allow ORCID users logged in the system to add OpenAIRE products to their ORCID Curriculum.

Acknowledgements This work is partly funded by the OpenAIRE-Advance H2020 project (grant number: 777541; call: H2020-EINFRA-2017) and the OpenAIRE-Connect H2020 project (grant number: 731011; call: H2020-EINFRA-2016-1). Moreover, we would like to thank our colleagues Michele Manunta, Francesco Casu, and Claudio De Luca (Institute for the Electromagnetic Sensing of the Environment, CNR, Italy) for their work on the EPOS infrastructure RCD; and Stephane Pesant (University of Bremen, Germany) his work on the European Marine Science RCD.

References

1. Artini, M., Atzori, C., Bardi, A., La Bruzzo, S., Manghi, P., Mannocci, A.: The openaire literature broker service for institutional repositories. *D-Lib Magazine* **21**(11/12), 1 (2015). <https://doi.org/10.1045/november2015-artini>

2. Atzori, C., Manghi, P., Bardi, A.: Gdup: De-duplication of scholarly communication big graphs. In: 2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT). pp. 142–151. IEEE (2018). <https://doi.org/10.1109/BDCAT.2018.00025>
3. Barker, M., Olabarriaga, S.D., Wilkins-Diehr, N., Gesing, S., Katz, D.S., Shand, S., Henwood, S., Glatard, T., Jeffery, K., Corrie, B., Treloar, A., Glaves, H., Wyborn, L., Hong, N.P.C., Costa, A.: The global impact of science gateways, virtual research environments and virtual laboratories. *Future Generation Computer Systems* **95**, 240 – 248 (2019). <https://doi.org/10.1016/j.future.2018.12.026>
4. Candela, L., Castelli, D., Pagano, P.: D4Science: an e-Infrastructure for Supporting Virtual Research Environments. In: Agosti, M., Esposito, F., Thanos, C. (eds.) *Post-proceedings of the Fifth Italian Research Conference on Digital Libraries - IRCDL 2009*, Padova, Italy, 29-30 January 2009. pp. 166–169. DELOS: an Association for Digital Libraries / Department of Information Engineering of the University of Padua (2009), <http://www.dei.unipd.it/%7Eagosti/ircdl/atti-ircdl2009-finale.pdf>
5. Candela, L., Castelli, D., Pagano, P.: Virtual research environments: an overview and a research agenda. *Data Science Journal* **12**, GRDI75–GRDI81 (2013). <https://doi.org/10.2481/dsj.GRDI-013>
6. De Roure, D.: The future of scholarly communications. *Insights* **27**(3) (2014). <https://doi.org/10.1629/2048-7754.171>
7. European Commission: Validation of the results of the public consultation on Science 2.0: Science in Transition. Tech. rep., European Commission (2015), http://ec.europa.eu/research/consultations/science-2.0/science_2.0_final_report.pdf
8. European Commission: Open Innovation, Open Science and Open to the World. Tech. rep., European Commission’s Directorate-General for Research & Innovation (2016), <https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe>
9. Grethe, J.S., Bandrowski, A., Banks, D.E., Condit, C., Gupta, A., Larson, S.D., Li, Y., Ozyurt, I.B., Stagg, A.M., Whetzel, P.L., Marengo, L., Miller, P., Wang, R., Shepherd, G.M., Martone, M.E.: SciCrunch: A cooperative and collaborative data and resource discovery platform for scientific communities. *Frontiers in Neuroinformatics* (69) (2014). <https://doi.org/10.3389/conf.fninf.2014.18.00069>
10. Jeffery, K.G., Meghini, C., Concordia, C., Patkos, T., Brasse, V., Ossenbruck, J.v., Marketakis, Y., Minadakis, N., Marchetti, E.: A reference architecture for virtual research environments. In: *Everything Changes, Everything Stays the Same? Understanding Information Spaces. Proceedings of the 15th International Symposium of Information Science (ISI 2017)*. pp. 76–88 (2017). <https://doi.org/10.18452/1448>
11. Kobos, M., Bolikowski, L., Horst, M., Manghi, P., Manola, N., Schirrwagen, J.: Information inference in scholarly communication infrastructures: the OpenAIREplus project experience. *Procedia Computer Science* **38**, 92–99 (2014). <https://doi.org/10.1016/j.procs.2014.10.016>
12. Manghi, P., Bardi, A., Atzori, C., Baglioni, M., Manola, N., Schirrwagen, J., Principe, P.: The OpenAIRE Research Graph Data Model (Apr 2019). <https://doi.org/10.5281/zenodo.2643199>
13. Manghi, P., Manunta, M., Baglioni, M., Bardi, A., Casu, F., De Luca, C., Kokogianaki, A.: Enabling Open Science publishing for Research Infrastructures via OpenAIRE: The EPOS use-case (Jul 2018). <https://doi.org/10.5281/zenodo.1412509>
14. Moore, R.L., Baru, C., Baxter, D., Fox, G.C., Majumdar, A., Papadopoulos, P., Pfeiffer, W., Sinkovits, R.S., Strande, S., Tatineni, M., Wagner, R.P., Wilkins-Diehr, N., Norman, M.L.: Gateways to discovery: Cyberinfrastructure for the long

tail of science. In: Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment. pp. 39:1–39:8. XSEDE '14, ACM, New York, NY, USA (2014). <https://doi.org/10.1145/2616498.2616540>