Open Access to Publications and Research Data in Horizon 2020: What Are the Requirements and How Can Institutional Repositories and OpenAIRE Help to Meet Them?

Eloy Rodrigues

University of Minho, Portuga
eloy@sdum.uminho.pt

1   Introduction

In the last decade the principle of Open Access to publicly funded research has been getting a growing support from policy makers and funders across Europe, both at national level and within the European Union context. At European level some of the first relevant steps taken by the European Research Council (ERC) with a statement supporting Open Access (2006), shortly followed by guidelines for researchers funded by the ERC (2007) stating that all peer-reviewed publications from ERC funded projects should be made openly accessible shortly after their publication. Those guidelines were revised in October 2013, reinforcing the mandatory character of the requirements and expanding them to monographs[1].

The European Commission took it first global political initiative in August 2008, launching an Open Access Pilot, covering seven areas (corresponding to about 20% of the total funding) of the 7th Framework Programme (FP7). All the grant agreements signed after August 2008 contained a clause (Special Clause 39) requiring beneficiaries to deposit articles resulting from FP7 projects into an repository and to make their best efforts to ensure open access to these articles within six months or twelve months.

Building on the results from those first initiatives, and on the growing political support to Open Access, the Commissioners Neelie Kroes (Vice-President of the European Commision and responsible for the Digital Agenda) and Máire Geoghegan-Quinn (Commisioner for Research, Innovation and Science), presented in July 2012, three fundamental documents[2], establishing Open Access as the "default" for the European Research Area, and the new Framework Programme, Horizon 2020.

---

[1] *Open Access Guidelines for researchers funded by the ERC* - http://erc.europa.eu/sites/default/files/document/file/ERC_Open_Access_Guidelines-revised_2013.pdf

[2] *Communication on a reinforced European Research Area partnership for excellence and growth* (http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-partnership-excellence-growth_en.pdf); *Communication Towards better access to scientific information: Boosting the benefits of public investments in research*

2   The Open Access Requirements in Horizon 2020

The underlying principle of Open Access to the research results funded in the Horizon 2020 announced in 2012, has already been defined and anchored on the documents and regulations of the new Framework Programme[3], as well as through the relevant provisions in the H2020 grant agreement[4]. As previously mentioned, the ERC has specific requirements, stated on their Open Access Guidelines for researchers funded by the ERC (see footnote 1), especially related with the locus of deposit. The requirements that will be detailed explained in this text are the ones from the general H2020 grant agreements, on not those from ERC grants.

Regarding publications, all beneficiaries of H2020 funding are required to deposit, as soon as possible and at the latest on publication, a copy of their peer reviewed journal articles into a repository, and must ensure open access to the deposited publication — via the repository — as soon as possible, but no later than six (or twelve for publications in the social sciences and humanities) months of publication.

Looking into those requirements[5] in a more detailed way, we can formulate and answer to several questions, namely:

What are the general Open Access requirements and who is covered by them?

All beneficiaries of H2020 funding must provide open access (free of charge, online access for any user) to all peer-reviewed publications by depositing them into a repository

---

    (http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf); *Recommendation on access to and preservation of scientific information* (http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf);

[3]   See *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020* (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf) and *Fact sheet: Open Access in Horizon 2020* (https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/FactSheet_Open_Access.pdf).

[4]   See Article 29 "Dissemination of results - Open Access - Visibility of EU funding" pages 60-63 of the *Multi-beneficiary General Model Grant Agreement*, Version 1.0, December 11, 2013 (http://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi_en.pdf).

[5]   See Article 29.2 " Open access to scientific publications" page 61 of the *Multi-beneficiary General Model Grant Agreement*, Version 1.0, December 11, 2013 (http://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi_en.pdf)

Where to deposit?

Researchers must deposit in a repository for scientific publications (online archives) of their choice. It can be an institutional repository (of the research institution with which they are affiliated) or a subject-based/thematic repository, or a centralized repository, like Zenodo[6].

What to deposit?

Researchers must deposit a machine-readable electronic copy of the published version. It may be the publisher's final version of the paper (including all modifications from the peer review process, copyediting and stylistic edits, and formatting changes), or the final peer-reviewed manuscript accepted for publication (final manuscript of a peer-reviewed paper accepted for journal publication, including all modifications from the peer review process, but not yet formatted by the publisher, also referred to as "post-print" version).

When to deposit?

Each beneficiary must deposit as soon as possible and at the latest on publication.

When should open access be provided?

Each beneficiary must ensure open access to the deposited publication — via the repository — as soon as possible but at the latest: (i) on publication, if an electronic version is available for free via the publisher, or (ii) within six months of publication (twelve months for publications in the social sciences and humanities) in any other case.

Are there metadata requirements on repository deposition?

Yes. Beneficiaries must ensure immediate open access to the bibliographic metadata that identify the deposited publication in a standard format and must include all of the following:

─ the terms ["European Union (EU)" and "Horizon 2020"]["Euratom" and Euratom research and training programme 2014-2018"];
─ the name of the action, acronym and grant number;
─ the publication date, and length of embargo period if applicable, and
─ a persistent identifier.

---

[6] Zenodo is a repository set up by the OpenAIRE project, available at http://www.zenodo.org/

Are Author Processing Charges (APCs) supported?

Yes. For open access publishing, researchers can publish in open access journals, or in journals that sell subscriptions and also offer the possibility of making individual articles openly accessible (hybrid journals). Where the case, the Author Processing Charges (APCs) incurred by beneficiaries are eligible for reimbursement during the duration of the action.

What are the consequences for non-compliance with the Open Access Requirements?

If a beneficiary breaches any of its obligations, the grant may be reduced (Article 43) and it may also lead to any of the other measures described in Chapter 6 of the General Model Grant Agreement.

On top of the above described requirements, the Commission encourages authors to retain their copyright and grant adequate licenses (like Creative Commons CC-BY licenses).

Beyond Open Access to publications, one of the novelties in Horizon 2020 is the Open Research Data Pilot which aims to improve and maximize access to and re-use of research data generated by EC funded projects. The results of this pilot will be monitored and will serve as step stone for developing new open research data policies in future Framework Programmes.

The participants of projects covered by the Open Research Data Pilot are required to deposit into a in a research data repository, and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate, the data needed to validate the results presented in scientific publications and other data specified in data management plan.

Examining those requirements[7] in a more specific way the following questions we will try to reply to the following questions:

What is the scope of the Open Research Data Pilot?

Projects participating in the following core areas of Horizon 2020 are part of the Open Research Data Pilot, namely:

— Future and Emerging Technologies
— Research infrastructures – part e-Infrastructures
— Leadership in enabling and industrial technologies – Information and Communication Technologies

---

7 See Article 29.3 " Open access to research data" page 61 of the *Multi-beneficiary General Model Grant Agreement*, Version 1.0, December 11, 2013 (http://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi_en.pdf)

- Societal Challenge: Secure, Clean and Efficient Energy – part Smart cities and communities
- Societal Challenge: Climate Action, Environment, Resource Efficiency and Raw materials – with the exception of raw materials topics
- Societal Challenge: Europe in a changing world – inclusive, innovative and reflective Societies
- Science with and for Society

Projects may opt out of the Pilot on Open Research Data in Horizon 2020 in a series of cases that include conflict with obligation to protect results, with confidentiality obligations, with security obligations or with rules on protection of personal data.

Projects funded under Horizon 2020 and not covered by the scope of the Pilot may participate on a voluntary basis ('opt in').

What are the requirements/what must be deposited?

Projects participating in the Pilot are required to deposit, and take measures to enable third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user:

- The data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;
- Other data, including associated metadata, as specified and within the deadlines laid down in a data management plan (DMP – see below).

At the same time, projects should provide information about tools and instruments at the disposal of the beneficiaries and necessary for validating the results, for instance specialised software or software code.

Where to deposit?

Projects should deposit in a research data repository. It could be a community/discipline specialized repository or Zenodo, provided by OpenAIRE.

What is a Data Management Plan?

A data management plan is a document outlining how research data will be handled during a research project, and after it is completed. A DMP describes what data will be collected / generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved.

Is there any support (financial or other) associated with the Open Research Data Pilot?

Costs relating to the implementation of the pilot will be reimbursed. Specific technical and professional support services will also be provided.

What are the consequences for non-compliance with the Open Research Data Pilot?

If a beneficiary breaches any of its obligations, the grant may be reduced (Article 43) and it may also lead to any of the other measures described in Chapter 6 of the General Model Grant Agreement.

3 OpenAIRE and Its Services for H2020 Compliance

OpenAIRE (Open Access Infrastructure for Research in Europe)[8] was originally a project established to support the policies of the EC's FP7 Open Access pilot, and the ERC's Open Access policy, but it has since then further developed into an Open Access infrastructure for Europe. Building on previous work on repository networking carried out by the DRIVER project[9], OpenAIRE has extended its gathering of open content to other data providers, such as national funders, current information systems (CRISs) and research data repositories.

Currently, OpenAIRE is a network of around 500 Open Access repositories, archives, journals and other data sources that support Open Access policies. It goes beyond the traditional publication aggregators by interconnecting entities related to scholarly communication (publications, research data, funding, people, organizations, data sources) allowing users to navigate alongside a rich information space graph and providing a wide range of services, from deposition to statistics.

Alongside with the technical e-infrastructure, the other fundamental building block of OpenAIRE's participatory design is a community-led network, involving all EC member states and 5 associate countries, which is a vital element for an open research infrastructure.

Through this network of 33 National Open Access Desks (NOADs) OpenAIRE engages multiple stakeholders across Europe: project coordinators and researchers, repository managers, research administrators at universities, and policy makers, to establish close collaboration and integrated workflows which make it easy for researchers and institutions to commit to and comply with the EC's, national and institutional open access policies.

The OpenAIRE e-infrastructure has been fully operational since December 2012, and its information is accessible both through the infrastructure portal and programmatically as a "service", through the OpenAIRE API[10]. Another key element of the

---

[8] www.openaire.eu

[9] http://www.driver-repository.eu/

[10] The documentation about the OpenAIRE API is available at http://api.openaire.eu/

success of OpenAIRE is its growing data capacity, i.e. the rich and varied set of data resources that it brings together and makes available to its potential users. As of August 2014, the OpenAIRE e-infrastructure provides access to more than 8 million publications (aggregated from almost five hundred data sources) with different modes of access, while the number of projects, mainly funded by the European Commission and the Wellcome Trust, is bigger than thirty thousand.

To achieve this impressive growth, OpenAIRE relies on its locally based communities of repository administrators as well as national and regional aggregation managers to carry out curatorial activities and to maintain datasets and registries. The infrastructure in turn exploits these locally created and maintained contents. Based on open protocols and agreements with the participating repositories, OpenAIRE retrieves their content (pdf) and apply a set of text mining algorithms to extract information, enrich the associated metadata and infer links among the various entities. More specifically OpenAIRE extract: (i) project information to link to funding (currently EC and Wellcome Trust, but expanding to other funders); (ii) titles, authors and their affiliations to organizations where possible; (iii) references to data and publications to build Open Access citation indexes.

The already mature, but yet evolving, OpenAIRE infrastructure provides several valuable services, to different stakeholders, which ensure and facilitate the compliance with the Open Access requirements set out for Horizon 2020. We will briefly mention here some of the most relevant ones for project participants (researchers and project managers) and for repository managers (willing to offer their repositories as compliance tool for their researchers).

Researchers and project managers are certainly one of the key stakeholders for OpenAIRE, which offers participants of Horizon 2020 funded project with a set of services and information to facilitate the compliance with their Open Access requirements. OpenAIRE provides researchers with detailed and practical information on how to comply with Horizon 2020 OA requirements[11] and offers them a tool to locate, and initiate, the deposit of their publications in the appropriate repositories (usually the repository of the research institution they are attached to)[12]. If project participants can't find an appropriate repository, they can deposit their publications publications and/or data into Zenodo, an OpenAIRE sponsored repository. Finally, OpenAIRE offers a service for authors to "claim" and link publications to funding (or data) when this is not automatically done by the OpenAIRE automatic processing.

Besides these services for depositing and linking research results, OpenAIRE also offers researchers and project coordinators a set of relevant services and tools for dissemination, reporting and monitoring.

The publications (automatically aggregated by the OpenAIRE portal from a compliant repository, or manually linked by researchers) which are correctly identified as produced by an Horizon 2020 project will be identified and presented in individual project pages. Each H2020 project has its own page on OpenAIRE, featuring project

---

[11] OpenAIRE Guide for Researchers - https://www.openaire.eu/researcher-toolkit/researcher-deposit-toolkit

[12] Available at https://www.openaire.eu/participate/deposit-publications-data

information, project publications and datasets, a statistics section and an "App box" with widgets and other tools.

The publications linked to the project and presented in the project page are reused and reusable, for dissemination, monitoring and reporting purposes in several ways:

— The publication list on OpenAIRE project page is automatically embedded in the EC's project portal (CORDIS);
— A publication list, for EC progress reporting, can be generated (in html format) or downloaded (in CSV format);
— The project list of publications can be incorporated into project Websites, or other external sites, through the widget (javascript) provided in the "App box".

Repository managers are another very important stakeholders for OpenAIRE, which also offers them a set of information and services to help the researchers from their institutions to comply with the H2020 requirements, by offering OpenAIRE compatible repositories, and other support services.

Besides a vast amount of information about Open Access, OpenAIRE provides repository managers with a detailed guide[13] on how to make their repositories compatible with the infrastructure by using the OpenAIRE Guidelines[14]. Repository managers can also use the OpenAIRE validator tool[15], an easy way to validate a repository (or journal) and register it into the OpenAIRE network.

Finally, OpenAIRE offers repository managers information about EC projects, to facilitate the linkage of publications to the associated projects during the deposit processes, and addons or plugins, to include that funding information, are available for some of the popular repository platforms.

---

[13] OpenAIRE Guide for Repository Managers at https://www.openaire.eu/rep-man-toolkit/repository-managers
[14] Available at https://guidelines.openaire.eu/wiki/Main_Page
[15] Available at http://validator.openaire.eu/