

European Commission Policy on Open-Access to Scientific Publications and Research Data in Horizon 2020

David Guedj (✉), **Celina Ramjoué**

*European Commission, Directorate General for Communications Networks, Content and Technology
Brussels, Belgium*

<http://ec.europa.eu/dgs/connect/en/content/dg-connect>

ARTICLE INFO:

RECEIVED: 09 Oct 2014
ACCEPTED: 10 Nov 2014
ONLINE: 02 Jan 2015

KEYWORDS:

Data management
Horizon 2020
Open Access
OA model
Open Research Data
Open Science

ABSTRACT

This paper briefly presents the main lines of the European Commission's policy on Open Access to Scientific Publications and Research Data in the Horizon 2020 programme. The respective pilot on open research data is presented in detail. The pilot covers approximately 20 percent of the H2020 funding in 2014 and 2015 in the areas of future and emerging technologies; e-Infrastructures; leadership in enabling and industrial technologies – ICT; societal challenges related to smart cities and communities, climate action, environment, and resource efficiency, inclusive, innovative and reflective societies; and science with and for society, thus extending the move from open access to research results towards OA to research data. A next step will be to consider a more global open science approach encompassing a change of culture towards more openness, accessibility, re-usability, multidisciplinary, participatory collaboration, and an approach driven by what is of benefit for society.



Creative Commons BY-NC-SA 4.0

© 2015 Hosting by Procon Ltd. All rights reserved.

Introduction

The European Commission has worked on open science-related issues since 2006, focusing first on open access to publications, later on open research data, and currently endeavouring to develop a vision for open science focusing on the transformation of science in the digital age. The building blocks of open science include open access to publications and data, citizen science and citizen participation in science, evidence-based policy making, alternative metrics for measuring the outputs of science and research, and e-infrastructures for open science.

Open Access policy in Horizon 2020

Open Access policy is part of the Commission's Europe 2020 strategy for a smart, sustainable and inclusive economy focusing on the key role of knowledge and innovation in generating growth. This strategy is implemented under the EU Framework Programme for Research and Innovation, Horizon 2020 (2014-2020). A fundamental principle underlying open access policy is that the research results of publicly funded European Research should be made freely accessible to all interested parties (i.e. researchers, industry and citizens). A wider and deeper access to scientific publications and data is conducive to improved quality

of scientific results, greater efficiency, faster innovation, and a greater transparency of the scientific process. These principles are translated into specific requirements in the Model Grant Agreement used to implement the Horizon 2020 Work Programme.

Definition of Open Access in Horizon 2020

Open-access is defined as the provision of on-line access to scientific information that is free of charge to the end-user and that is, as far as possible, re-usable. Key elements of open access in the case of publications include the right to read, download and print, the right to copy, distribute, search, link, crawl and mine. Openly accessible research data can typically be accessed, mined, exploited, reproduced and disseminated free of charge for the user.

In Horizon 2020, “scientific information” refers essentially to peer-reviewed articles published in scholarly journals and research data, including data underlying publications, curated data and/or raw data. The two main routes to OA are:

- self-archiving (“Green” open access)
- open-access publishing (“Gold” open access).

Green open-access

“Green open access” means that the published article or the final peer-reviewed manuscript is archived (deposited) by the author—or a representative—in an online repository before, alongside or after its publication. Repository software usually allows authors to delay access to the article if required by the publisher (“embargo period”).

Gold open-access

A “Gold open access” article is immediately provided in open access mode. In this model, the payment of publication costs is shifted away from readers paying via subscriptions. The business model most often encountered is based on one-off payments by authors. These costs (often referred to as Author Processing Charges, APCs) can sometimes be borne by the university or research institute to which the author is affiliated, or by the funding agency supporting the research. In other cases, the costs of open access publishing are covered by subsidies or other funding models.

Open access does not imply that there is an obligation to publish results. Researchers are entirely free to publish or not. Therefore, Open Access does not interfere with the decision to exploit research results commercially.

Mandate on Open Access to Publications

Under Horizon 2020, each beneficiary must ensure open access to all peer-reviewed scientific publications relating to its results. In concrete terms, beneficiaries must ensure that their publications can be read online, downloaded and printed. Any other means to increase usage of the publications (e.g. copy, distribute, search, link, crawl, mine, etc.) should also be sought.

The Open Access mandate consists of two steps:

Step 1: Beneficiaries must deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications, of their choice. This should be done as soon as possible and definitely before publication. The format of the publication must be usable and understandable by a computer, using well-known or standardly used text file formats. A repository for scientific publications is an online archive. The Open Access Infrastructure for Research in Europe (OpenAIRE¹) is the recommended entry point for researchers to determine what repository to choose. At the same time, the beneficiary must strive to deposit the supporting research data, ideally into a data repository.²

Step 2: After depositing publications, and whenever possible – underlying data, beneficiaries must ensure open access to the deposited publication and data via the chosen repository.

Beneficiaries can choose either the “Green Open-Access” model or the “Gold Open-Access” model. In the case of Green OA, access must be granted after an embargo period of no more than six months. In the case of Gold OA, APCs are eligible for reimbursement if incurred during project duration.

Metadata should also be made open access. Metadata is important since it helps to identify the deposited publications, access them and acknowledge EU funding in the case of point.

In all cases, the Commission encourages authors to retain their copyright and grant adequate licenses to publishers. A useful licensing solution is offered through Creative Commons (e.g. CC-BY or CC-0 licenses).³

Where possible, it is also recommended that contributors be uniquely identifiable, and data uniquely attributable, through identifiers which are persistent, non-proprietary, open and interoperable (e.g. through leveraging existing sustainable initiatives such as ORCID for contributor identifiers and DataCite for data identifiers).

Horizon 2020 Pilot on Open Research Data

The Open Research Data Pilot aims to improve and maximize access to and re-use of research data generated by projects. When used in connection with research data, “open access” typically refers to research data that can be accessed, mined, exploited, reproduced and disseminated free of charge for the user. The legal requirements for projects participating in this pilot are contained in the optional article 29.3 of the Model Grant Agreement.

Scope of the Pilot

The scope of the Pilot for 2014-2015 comprises the following areas:

- Future and Emerging Technologies
- Research Infrastructures – part e-Infrastructures
- Leadership in enabling and industrial technologies – Information and Communication Technologies
- Societal Challenge: “Secure, Clean and Efficient Energy” – part Smart cities and communities
- Societal Challenge: “Climate Action, Environment, Resource Efficiency and Raw materials” – except raw materials
- Societal Challenge: “Europe in a changing world – Inclusive, Innovative and Reflective Societies”
- Science with and for Society.

This corresponds to about €3 billion or 20 percent of the overall Horizon 2020 budget in 2014 and 2015.

It is important to note that voluntary participation in the Pilot is also possible.

The pilot covers the following types of data:

1) data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible.

2) other data, including associated metadata, as specified and within the deadlines laid down in the data management plan (DMP).

A DMP is a document outlining how the research data collected or generated will be handled during the lifetime of a project, and after it is completed, describing 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.⁴ A detailed data management plan, covering individual datasets, is required for funded projects participating in the Open Research Data Pilot, with a first version of the DMP to be provided within six months from

the project start. For other funded projects, a DMP is not required although they may deem fit to submit one. Data management is of key importance in the Pilot on Open Research Data. One of its main objectives is to raise awareness on research data management.

Provisions of the Open Research Data Pilot

Beneficiaries participating in the Pilot are required to:

- Deposit a) underlying data and b) “other data” as specified in the Data Management Plan (DMP) into a research data repository of their choice;
- Take measures to make it possible to access, mine, exploit, reproduce and disseminate free of charge (using e.g. Creative Commons licenses);
- Provide information about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (where possible, provide the tools and instruments themselves).

It is important to note that beneficiaries participating in the Pilot are not obliged to make all datasets open (as described in their DMP, and in compliance with confidentiality, security, data protection, and other relevant considerations).

Total or partial opting out of the Pilot Action on Open Research Data

Projects may opt out of the Pilot at submission stage for a variety of reasons, including an obligation to protect results, for confidentiality or security reasons, in order to ensure the protection of personal data, to guarantee the achievement of the main aim of the action, or if no research data is generated/ collected by the project. Information provided on opting for the Pilot is not part of the evaluation of the proposal, and therefore has no influence on the evaluation of the proposal itself.

Conclusion

Promoting open access to scientific publications at European level has had the positive effect of encouraging other funding bodies and institutions across Europe to develop their own open access policies. In this respect, the work undertaken directly with representatives of the Member States (National Points of Reference on access to and preservation of scientific information) has been valuable. Researchers across Europe are increasingly used to the idea of sharing their research results. In the digital era, an open question remains

whether the right to read is also a right to mine scientific publications. This question is linked to the ongoing debate on copyright, and, specifically, text- and data-mining.

The current step, as exemplified by the Open Research Data Pilot, is to extend the move towards open access to research data. Whereas most Member States within the European Research Area have declared that they had OA policies for scholarly publications, it is much less the case for OA for scientific data. Some reasons are the reluctance of researchers and the resistance of specific stakeholders, e.g. industry. Other reasons have to do with the lack of coordination between relevant databases that prevents an effective exchange of data between organisations that have a common interest in accessing this data. In order to overcome these barriers or resistances, there is a need to show the benefits of opening up research data, including for organisations such as governmental agencies, funding agencies and others.

One of the strongest incentives for the generalization of open research data is the ability for the scientific community to find and access data easily, free of charge and re-use it either to reproduce experiments or to discover new properties and allow scientific breakthroughs. As for peer-review for scientific publications, applying this process to research data would reinforce the credibility and reputation of the authors or conversely detect anomalies or frauds. It could also serve as a career booster by applying alternative metrics to those that are currently in use and that suffer from serious limitations. More emphasis could for example be put on the amount and quality of collaborations rather than on the citations in high-impact scientific journals. In addition, some scientific communities, like genetics or molecular biology, have been used for a long time now to share their research data and collaborate, leading to scientific breakthroughs or at least acceleration in the achievement of major scientific results. Further, open data promotes economic development

by the products and services that are derived from it and through the extensive transformation that adds value to the original information.

Finally, taking a broader perspective, a next step will be to consider a more global open science approach encompassing a change of culture towards more openness, accessibility, re-usability, multidisciplinary, participatory collaboration, and an approach driven by what is of benefit for society.

Disclaimer

The opinions expressed are those of the authors only and should not be considered as representative of the European Commission's official position.

References

¹ OpenAIRE, a three-year project, will establish the infrastructure for researchers to support them in complying with the EC OA pilot and the ERC Guidelines on Open Access. It will provide an extensive European Helpdesk System, based on a distributed network of national and regional liaison offices in 27 countries, to ensure localized help to researchers within their own context. It will build an OpenAIRE portal and e-Infrastructure for the repository networks and explore scientific data management services together with 5 disciplinary communities. It will also provide a repository facility for researchers who do not have access to an institutional or discipline-specific repository.

² More information on registries of data repositories at <http://www.re3data.org/> and <http://zenodo.org/>.

³ See Creative Commons, About the Licenses, Available from: <http://creativecommons.org/licenses/>.

⁴ See separate *Guidelines on Data Management in Horizon 2020*, version 1.0, 11 December 2013, available at http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.