Open Science and Horizon 2020

The article provides an introduction to open science, highlights open access to publications mandate and Open Research Data Pilot in the Horizon 2020 and features open reproducible projects and tools.

What is open science and why is it important?

Open science is the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society, amateur or professional [1]. Open science calls for more transparency in experimental methodology, observation, and collection of data; public availability, accessibility, reusability and transparency of research data and scholarly communication; and encourages the use of web-based tools to facilitate scientific collaboration [2]. Open access to scientific publications and optimal reuse of research data are the most important aspects of open science, but it also includes citizen science, and research integrity and reproducibility.

There is a general consensus that the results of publicly funded research should be made available in an as open as possible manner. Unnecessary legal, organisational and financial barriers to access publicly funded research results should be removed (as much as possible and appropriate) to maximise knowledge sharing, taking into account (when necessary) the need for exploitation of results.

Open science has the potential to increase the quality, impact and benefits of science and to accelerate advancement of knowledge by making it more reliable, more efficient and accurate, better understandable by society and responsive to societal challenges, and has the potential to enable growth and innovation through reuse of scientific results by all stakeholders at all levels of society, and ultimately contribute to growth and competitiveness of the country and the region.

Open access policies in Horizon 2020

Open Science is one of the three pillars of the European Commission Vision for Europe [3] together with Open Innovation and Open to the World.

It is now widely recognised that making research results more accessible to all societal actors contributes to better and more efficient science, and to innovation in the public and private sectors. The European Commission therefore supports open access at the European level (in its Research Framework Programmes), at the Member States level and internationally:

- **Peer-reviewed scientific publications** (primarily research articles published in academic journals): All projects receiving Horizon 2020 funding are **required** to make sure that each peer-reviewed journal article they publish is openly accessible, free of charge (article 29.2. Model Grant Agreement).
- **Scientific research data**: data underlying publications and/or other data (such as curated but unpublished datasets or raw data): The European Commission is running a **pilot on open access to research data** in Horizon 2020: the Open Research Data (ORD) pilot. [4]

Open access to publications mandate

All projects receiving Horizon 2020 funding are required to make sure that the peer reviewed publications resulting from the project are available in open access. A machine-readable electronic copy of the published version (publisher's final version of the paper, usually a PDF document) or a final peer-reviewed manuscript accepted for publication (including all modifications from the peer review process, but not yet formatted by the publisher, also referred to as "post-print" version) should be deposited in a repository for scientific publications (institutional, subject-based/thematic or Zenodo [5] as soon as possible and at the latest on publication and open access should be ensured on publication, if an electronic version is available for free via the publisher or within six months of publication (12 months for publications in the social sciences and humanities) in any other case. Publishing in open access journals is also welcomed and Article Processing Charges (where the case) are eligible costs within the project budget [6].

The Open Research Data Pilot: "As open as possible, as closed as necessary"

Open data is data that is free to access, reuse, repurpose, and redistribute. The Open Research Data Pilot aims to make the research data generated by Horizon 2020 projects accessible with as few restrictions as possible, while at the same time protecting sensitive data from inappropriate access.

Horizon 2020 projects must deposit their data in a research data repository where they will be findable and accessible for others. Do not panic, there is no requirement to share sensitive data or breach any Intellectual Property Rights agreements with industrial partners. And there is no pressure to deposit all the data that the projects generate during the project either – only the data, which underpins published research findings and/or has a longer-term value. In addition to supporting research integrity, openness has many other benefits. Improved visibility means the research will reach more people and have a greater impact – for science, society and researchers careers. Recent studies have shown that citations increase when data is made available alongside the publication [7]; these papers also have a longer shelf-life.

How to make data open? Choose your dataset(s), apply an open content license, e.g. Creative Commons Attribution (CC-BY) [8], make the data available and discoverable via data repositories [9]. It is a part of good research practice, cuts down on

academic fraud, helps in validation of research results and leads to more scientific breakthroughs, the progress on Alzheimer's research is just one example of many.

To optimise the potential for future sharing and reuse, a Data Management Plan (DMP) will guide you through the process, help to consider any problems or challenges that may be encountered and identify ways to overcome these. A DMP should be thought of as a "living" document outlining how the research data collected or generated will be handled during and after a research project. Horizon 2020 projects have to deliver the first version of the DMP is within the first six months of the project, but there is no pressure to provide detailed answers to all the questions yet. The DMP describes the datasets, standards and metadata, data sharing, archiving and preservation. DMPonline [10], developed by the Digital Curation Centre and the University of California Curation Center, helps to create, review, and share data management plans that meet institutional and funder requirements. DMP needs to be updated over the course of the project whenever significant changes arise, such as new data or changes in the consortium policies or consortium composition [11].

Open reproducible science: projects and tools

"Mostly due to current methods capture and data malpractice, approximately 50 % of all research data and experiments is considered not reproducible, and the vast majority (likely over 80 %) of data never makes it to a trusted and sustainable repository", says the Commission High Level Expert Group on the European Open Science Cloud in their report "Realising the European Open Science Cloud" [12].

"Scientists like to think of science as self-correcting. To an alarming degree, it is not" was the highlight of the Economist article "Unreliable research: Trouble at the lab" addressing the reproducibility issues [13].

A number of open reproducible science projects has been addressing these issues. For example, Cancer Biology Reproducibility Project – a collaboration between Science Exchange and the Center for Open Science independently replicating a subset of experimental results from a number of high-profile papers in the field of cancer biology published between 2010 – 2012 – published a series of reviews underling the mixed results they received and demonstrating how difficult replication of published research results could be [14].

Preregistration

There are many innovations to improve the rigor and credibility of research. One of those is preregistration — making commitments to the design and analysis plan in advance of knowing the outcomes. Preregistration separates hypothesis-generating (exploratory) from hypothesis-testing (confirmatory) research. Both are important, but the same data cannot be used to generate and test a hypothesis, which can happen unintentionally and reduce the clarity and quality of results. Removing these potential conflicts through planning improves the quality and transparency of research, helping others who may wish to build on it.

The Center for Open Science runs the Preregistration Challenge providing an easy to use form for creating a rigorous research plan and awarding $1,000 prizes for one thousand researchers who publish the results of their preregistered work [15].

AsPredicted is another tool that could be used to pre-register research [16].

Sharing methods and materials, notebooks and codes

Open reproducible science also includes sharing methods and materials, e.g. via protocols.io or or MyExperiment [17]; sharing notebooks, e.g. at ONSNetwork [18] and code e.g. at GitHub [19]with GNU [20] or MIT license [21].

To create and share documents that contain live code, equations, visualizations and explanatory text; to clean and transform data, run numerical simulation, statistical modelling, machine learning; write research papers and much more researchers can use another open-source web application – the Jupyter Notebook [22].

Preprints

A number of scientist-driven initiatives, e.g. ASAPbio [23], promote the productive use of preprints. A preprint is a complete scientific manuscript that is uploaded by the authors to a public repository. The preprint contains complete data and methodologies and it is often the same manuscript being submitted to a journal. After a brief quality-control inspection to ensure that the work is scientific in nature, the author's manuscript is posted within a day or so on the preprints repository and can be viewed without charge by anyone in the world. Based upon feedback and/or new data, new versions of preprint can be submitted; however, prior preprint versions are also retained. Preprint repositories allow researchers to directly control the dissemination of their work to the world-wide scientific community. In most cases, the same work posted as preprint also is submitted for peer review at a journal. Thus, preprints (rapid, but not validated through peer-review) and journal publication (slow, but providing validation using peer-review) work in parallel as a communication system for scientific research.

References:

[1]   Open Science: *https://en.wikipedia.org/wiki/Open_science*

[2]   [The Open Science Project, What exactly is open science: *http://www.openscience.org/blog/?p=269*

[3]   European Commission: Research & Innovation: Open Science: *http://ec.europa.eu/research/openscience/index.cfm?pg=home*

[4]   European Commission: Research & Innovation: Open Science: Open Access: *https://ec.europa.eu/research/opescience/index.cfm?pg=openaccess*

[5]   Zenodo: *https://zenodo.org*

[6]   Open Access in Horizon 2020: *https://www.openaire.eu/open-access-in-horizon-2020*

[7]   Piwowar, H. & Vision, T.: Data reuse and the open data citation advantage: *https://peerj.com/articles/175*

[8]   Creative Commons Attribution 4.0 International licence (CC BY 4.0): *https://creativecommons.org/licenses/by/4.0*

[9]   Re3data.org (Registry of Research Data Repositories): *http://www.re3data.org*

[10] DMPonline: *https://dmponline.dcc.ac.uk*

[11] What is the Open Research Data Pilot: *https://www.openaire.eu/opendatapilot*

[12] Realising the European Open Science Cloud, EC DG Research & Innovation 2016 *http://ec.europa.eu/research/ openscience/pdf/realising_the_european_open_science_cloud_2016.pdf#view=fit&pagemode=none*

[13] Unreliable research: Trouble at the lab: *https://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble*

[14] The Reproducibility Project: Cancer Biology *https://osf.io/e81xl/wiki/home*

[15] Preregistration Challenge: *https://cos.io/prereg*

[16] AsPredicted: *https://aspredicted.org*

[17] MyExperiment *https://www.myexperiment.org*

[18] ONSNetwork (Open Notebook Science Network): *http://onsnetwork.org*

[19] GitHub (software development platform): *https://github.com*

[20] GNU General Public License: *https://www.gnu.org/licenses/gpl-3.0.en.html*

[21] MIT license: *https://en.wikipedia.org/wiki/MIT_License*

[22] The Jupyter Notebook: *http://jupyter.org*

[23] ASAPbio | Accelerating Science and Publication in Biology *http://asapbio.org; http://asapbio.org/preprint-info*

Gwen Franck   **Iryna Kuchma**

gwen.franck@eifl.net   iryna.kuchma@eifl.net   ■

(EIFL – Electronic Information for Libraries)