Open Data, Open Science, Open Access

Presentation by Sara Di Giorgio,

Crete, May 2017

1

The use of Open Data and Open Access is an integral element of Open Science. Like an astronaut on Mars, we're all engaged to explore key aspects and challenges of the currently transforming science landscape.

Open Science describes the on-going evolution in the modus operandi of doing research and organising science. This evolution is enabled by Big Data and digital technologies and is driven by both the globalisation of the scientific community and increasing public demand to address the societal challenges of our times. Open Science has, as a logical consequence, an ongoing transition in the way research is performed, researchers collaborate, knowledge is shared, and science is organised.

Open Science has an impact on the entire life cycle of research, from the starting point of research to its publication, and on how this cycle is organised.

Thus Open Science means openness of research data, methods, results and publications within the limits of research agreements.

To develop Open Science there is a need to focus on open data, open access and research integrity.

2

The trend towards Open Access is redefining the framework conditions for science and thus has an impact on how (Open) innovation is produced by encouraging a more dynamic circulation of knowledge.

The European Commission's policy actions on Open Science aim to strengthen knowledge circulation, since Open Science is, in practice, about sharing knowledge as early as practically possible in the discovery process.

Open Science is one of three priority areas for European research, science and innovation policy. The policy was presented by Carlos Moedas – Commissioner for Research, Science and Innovation in Brussels, in a speech titled 'A new start for Europe: Opening up to an ERA of Innovation' at the 'Open Innovation, Open Science, Open to the World' Conference in 2015.
"Openness will define our century", Commissioner Carlos Moedas, also says.

Because Openness is a basic principle in science, but supported by digital innovation, it can enable an increase of scientific production, a remote collaboration of scientists with the use of the web and more promising start-ups to emerge thanks to the exploitation of openly accessible research results.

"Open Science depends on open minds", according to Neelie Kroes, the former European commissioner.

However, Open Science does not mean 'free science'. It is essential to ensure that intellectual property is protected before making knowledge publicly available in order to attract investments that can help translate research results into innovation. If this is taken into account, fuller and wider access to scientific publications and research data can help to accelerate innovation.

Sharing information & enabling networking is the key to approaching our future research needs.

Open Science Monitor, an initiative of the European Commission to keep up-to-date with **progress in open science**, explores open science characteristics and indicators, where you can see that indicators are for both open access to publications and open scholarly communication.

3

"Science carried out and communicated in a manner which allows others to contribute, collaborate and add to the research effort, with all kinds of data, results and protocols made freely available at different stages of the research process". This is the definition of the Research Information Network in the UK that captures a systemic change to the way science and research have been carried out for the last fifty years: shifting from the standard practices of publishing research results in scientific publications towards sharing and using all available knowledge at an earlier stage in the research process.

So Open Science is more than open access publishing as you can see in the diagram of Andreas Neuhold from the DCC. It includes Open Data, Open Source that means sharing code and tools to allow others to reproduce work and Open methods, for Documenting and sharing workflows and methods and using web based tools to facilitate collaboration and interaction from the outside world

Open Science is changing the whole 'business cycle' of doing science and research – from the selection of research subjects, to the carrying out of research and to its use and re-use - as well as all the actors and actions involved. Such as universities and publishers.

Scientists, citizens, publishers, research institutions, public and private research funders, students and education professionals as well as companies from around the globe are sharing an open, virtual environment.

4

Research has a long history in Europe. Over recent decades, it has gradually acquired the sense of deliberate collaboration between European countries linking first their research activities, then their policies in this field.

The very first European-wide Framework Programme for research was launched 30 years ago to bring together expertise from across the European Community, as it was then known, and make Europe more competitive in key technologies. European research funding evolved from separate programmes to become a major component of the research and innovation landscape in Europe.

The Framework Programmes have become a major part of research cooperation in Europe. Their objective has also evolved from supporting cross-border collaboration in research and technology to now encouraging a European coordination of activities and policies. The reason for this is simple: research, technology and innovation are at the core of Europe's economy and are vital for a successful society.

Today, Horizon 2020, the eighth Framework Programme, is the biggest and most ambitious with a budget of EUR 80 billion.

With Horizon 2020, research and innovation will play a vital role in European Commission

5

From Open Science to European Science Cloud: the Commission plans to create a new **European Open Science Cloud** that will offer Europe's researchers a virtual environment to store, share and re-use their data across disciplines and borders, by integrating and consolidating e-infrastructure platforms, federating existing scientific clouds and research infrastructures, and supporting the development of cloud-based services.

This will be underpinned by the **European Data Infrastructure**, deploying the high-bandwidth networks, large-scale storage facilities and super-computer capacity necessary to effectively access and process large datasets stored in the cloud. This world-class infrastructure will ensure Europe participates in the global race for high performance computing in line with its economic and knowledge potential.

Focusing initially on the scientific community the user base will over time be enlarged to the public sector and to industry. This initiative is part of a package of measures to strengthen Europe's position in data-driven innovation, to improve competitiveness and cohesion and to help create a Digital Single Market in Europe.

The European Cloud Initiative will unlock the value of big data by providing world-class supercomputing capability, high-speed connectivity and leading-edge data and software services for science, industry and the public sector *by 2020.*

This initiative will make science more efficient and productive and let millions of researchers share and analyse research data in a trusted environment across technologies, disciplines and borders.

The European Cloud Initiative will make it easier for researchers and innovators to access and re-use data, and will reduce the cost of data storage and high-performance analysis. Making research data openly available can help boost Europe's competitiveness by benefitting start-ups, SMEs and data-driven innovation.

6

The Open Definition sets out principles that define “openness” in relation to data and content. “Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)”.

Availability and access: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.

Reuse and redistribution: the data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets. The data must be machine-readable.

Universal participation: everyone must be able to use, reuse and redistribute — there should be no discrimination against fields of endeavor or against persons or groups. For example, ‘non-commercial’ restrictions that would prevent ‘commercial’ use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

What kinds of open data?
There are many kinds of open data that have potential uses and applications:
Culture: Data about cultural works and artifacts — for example titles and authors — and generally collected and held by galleries, libraries, archives and museums.
Science: Data that is produced as part of scientific research
Finance: Data such as government accounts (expenditure and revenue) and information on financial markets (stocks, shares, bonds etc).
Statistics: Data produced by statistical offices such as the census and key socioeconomic indicators.
Weather: The many types of information used to understand and predict the weather and climate.
Environment: Information related to the natural environment such presence and level of pollutants, the quality and rivers and seas.

These are the most adopted licences for open data.

Explicit dedication of data underlying published science into the public domain via PDDL or CC0 is strongly recommended and ensures compliance with both the Science Commons Protocol for Implementing Open Access Data and the Open Knowledge/Data Definition.

CC-BY because data must credit whoever has published or generate the data (attribution) copies or adaptations of the data must be released similarly as open data (share-alike).

7

To be Open Data should be Available and FAIR, according to the initiative developed by Force 11 and now adopted by EC for H2020 project.
To be **Findable**:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

F2. data are described with rich metadata.
F3. (meta)data are registered or indexed in a searchable resource.
F4. metadata specify the data identifier.

To be **Accessible**:

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2 metadata are accessible, even when the data are no longer available.

To be **Interoperable**:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

To be **Re-usable**:

R1. meta(data) have a plurality of accurate and relevant attributes.
R1.1. (meta)data are released with a clear and accessible data usage license.
R1.2. (meta)data are associated with their provenance.
R1.3. (meta)data meet domain-relevant community standards.

8

How to make open data? You can follow the guidelines of the Open knowledge Foundation to help you in the process.

1. Choose your dataset(s)

- What can you may open? You may need to revisit this step if you encounter problems later. They recommend that you ask the community in the first instance. That is the people who will be accessing and using the data, as they are likely to have a good understanding of which data could be valuable.

2. Apply an open license

- Determine what IP exists. Apply a suitable licence e.g. CC-BY. Even in places where the existence of rights is uncertain, it is important to apply a license simply for the sake of clarity.

3. Make the data available

Provide the data in a suitable format. Use relevant metadata standard Use repositories. Open data needs to be technically open as well as legally open. Specifically, the data needs to be available in bulk in a machine-readable format. To make sure data can be understood by their community and others, researchers should create metadata and documentation. Metadata is basic descriptive information to help identify and understand the structure of the data e.g. title, author. Documentation provides the wider context. It's useful to share the methodology / workflow, software and any information needed to understand the data e.g. explanation of abbreviations or acronyms

4. Make it discoverable

There are lots of standards that can be used. The PARTHENOS started a catalogue of disciplinary metadata standards

9

5. Use appropriate file formats.

If you want your data to be re-used and sustainable in the long-term, you need to use open, non-proprietary formats.

10

Open Access (OA) means free and online access to scholarly literature, free of charge, and free of most copyright and licensing restrictions **that can be freely disseminated further with proper author attribution.**

What makes it possible is the internet and the consent of the author or copyright-holder. The internet has provided an alternative means of peer review, publication and distribution, and the on-line scholarly communication advances the science without loss of time

Self-archiving / 'green' open access – the author, or a representative, archives (deposits) the published article or the final peer-reviewed manuscript in an online repository before, at the same time as, or after publication. Some publishers request that open access be granted only after an embargo period has elapsed.

Open access publishing / 'gold' open access - an article is immediately published in open access mode. In this model, the payment of publication costs is shifted away from subscribing readers. The most common business model is based on one-off payments by authors. These costs, often referred to as Article Processing Charges (APCs) are usually borne by the researcher's university or research institute or the agency funding the research. In other cases, the costs of open access publishing are covered by subsidies or other funding models.

11

Benefits of open access are:

- Researchers can access relevant literature – not behind pay walls
- It ensures research is transparent and reproducible
- Increased visibility, usage and impact of your work
- Foster new collaborations and research partnerships
- Ensure long-term access to your outputs
- Help increase the efficiency of research

Certain research communities have also seen the benefit of sharing data as it speeds up the process of discovery. Researchers in the field of Alzheimer's research have agreed as a community to share data immediately to make scientific breakthroughs.

There's also a citation advantage for individual researchers. It's now clear that shared data received 9% more citations. *Articles freely available online are more highly cited.*

11

If you want to learn more have a look to whyopenreseach.org you will find statistics, information on opening up research, making it accessible and reusable by all. We view access to information as a human right and think it should be treated as such. We extend our thanks to the Shuttleworth Foundation.

12

As a funder, the European Commission is pushing to Open Access … WHY HAVE OPEN ACCESS TO PUBLICATIONS AND DATA IN HORIZON 2020?

- Build on previous research results (improved quality of results)
- Encourage collaboration and avoid duplication of effort (greater efficiency)
- Speed up innovation (faster progress to market means faster growth)
- Involve citizens and society (improved transparency of the scientific process).

There is a consideration made by the EC as founder "The European Commission's vision is that information already paid for by the public purse should not be paid for again each time it is accessed or used, and that it should benefit European companies and citizens to the full."

13

Carlos Moedas, European Commissioner for Research, Science and Innovation, launched the challenge for Europe Research: We need more open access to research results and the underlying data. Open access publication is already a requirement under Horizon 2020, but we now need to look seriously at open data […] Let's dare to make Europe open to innovation, open to science and open to the world".

And we can keep this as conclusion of this presentation, because most of the data today is locked in literature, is locked in papers, and we have to unlock that data.

You will find A Vision for Europe in "**The Three Os – Open Innovation, Open Science, Open to the World a European Open Science Cloud"** (see 'Further Reading' in this section)