

Some Thoughts about Open Science: A Few Challenges and a Great Opportunity for Science

Cristina Pujades
Universitat Pompeu Fabra

Algunas reflexiones sobre la ciencia abierta: retos y una gran oportunidad para la ciencia

ABSTRACT RESUMEN

Open Science focuses on spreading knowledge generated by scientific research as soon as it is available using digital and collaborative technology. The aim is to make science more accessible, inclusive, and equitable for the benefit of all. In this article, I will discuss some of the challenges and opportunities of this approach to the scientific process towards a comprehensive, effective open science from my perspective as a scientist in biology and health sciences. My aim is to share this with the academic community, since as Vice-Rector for Research and as scientist, I think this is one opportunity that we do not want to miss.

La Ciencia Abierta se basa en difundir el conocimiento generado por la investigación científica tan pronto como esté disponible mediante tecnología digital y colaborativa. El objetivo es hacer que la ciencia sea más accesible, inclusiva y equitativa en beneficio de todos. En este artículo, hablaré de algunos de los retos y oportunidades de esta visión del proceso científico hacia una ciencia abierta amplia y eficaz desde mi perspectiva como científica en biología y ciencias de la salud. Mi objetivo es compartirla con la comunidad académica, ya que, como Vicerrectora de Investigación y como científica, pienso que esta es una de las oportunidades que no queremos perder.

KEYWORDS PALABRAS CLAVE

FAIR principles; Open Science guidelines; Open Access publications; Scientific careers; Academic evaluations.

Principios FAIR; Directrices de Ciencia Abierta; Publicaciones en Acceso Abierto; Carreras científicas; Evaluaciones académicas.

Algunes reflexions sobre la ciència oberta: reptes i una gran oportunitat per la ciència

RESUM

La Ciència Oberta es basa en difondre el coneixement generat per la investigació científica tan bon punt estigui disponible mitjançant tecnologia digital i col·laborativa. L'objectiu és fer que la ciència sigui més accessible, inclusiva i equitativa en benefici de tots. En aquest article, parlaré d'alguns dels reptes i oportunitats d'aquesta visió del procés científic cap a una ciència oberta àmplia i eficaç des de la meua perspectiva com a científica en biologia i ciències de la salut. El meu objectiu és compartir-la amb la comunitat acadèmica, ja que com Vicerectora de Recerca i com a científica, penso que aquesta és una de les oportunitats que no volem perdre.

PARAULES CLAU

Principis FAIR; Directrius de Ciència Oberta; Publicacions en Accés Obert; Carreres científiques; Avaluacions acadèmiques.

1. Introduction

Open Science is an approach to the scientific process that focuses on spreading knowledge generated by scientific research (including publications, data, samples, and software) as soon as it is available using digital and collaborative technology. The aim, as stated by UNESCO, is to make science more accessible, inclusive and equitable for the benefit of all. By broadening the principles of openness to the whole research cycle, Open Science fosters sharing and collaboration, and very important, a change in the current scientific publishing model. I am sure most of us agree with this aim and we are eager to follow it. However, the transition towards a comprehensive, effective open science is not easy and requires a significant cultural change to ensure that scientific efforts have a real-world impact.

2. Data and software sharing

Already in 2016, a diverse set of stakeholders —representing academia, industry, funding agencies, and scholarly publishers— demanded guiding principles for the management and stewardship of data. They elaborated the FAIR Principles (Jacobsen et al., 2020; Wilkinson et al., 2016), which improve Findability, Accessibility, Interoperability, and Reuse of digital assets. They put specific emphasis on enhancing the ability of machines to automatically find and use the data, because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and speed of data creation. Nowadays, there are many available public repositories to share data, samples and software, and all of them involve the following of an ethic procedure. Institutions are increasing their capacity in Text and Data Mining practices —technical and legal— providing help to their scientists to follow the FAIR principles. However, there are still some challenges to face.

The data-sharing within the scientific community has been a common practice since many years. In my opinion as biologist, a turning point in collaborative/open science was the development of the publicly funded Human Genome Project (Consortium et al., 2001). It was launched in 1990 and accomplished in 2003, and remains the world's largest collaborative biological project. The Human Genome Project produced the first draft of the human genome sequence, which has been followed by the sequencing of many more genomes. It became a crucial moment in which scientists realized that a collaboration was more than the sum of different parts. Several browsers were developed to help and boost the study of genes and genomes. The reference one, Ensembl¹, allows to find any of the available sequenced genomes (from *Caenorhabditis elegans* to humans). This collaborative effort was followed by several other examples of Open Science projects in biology, which focus on gathering vast collections of data.

For instance, we find the Encyclopaedia of DNA elements Consortium, ENCODE², which aims to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control when and where a gene is active. Or the Encyclopaedia of Life (EOL, Smithsonian³), an online encyclopaedia intended to document all living species to provide global access to knowledge about life on Earth. All these projects started long ago, and their commonalities are that they are successful in generating new knowledge, they produce similar types of datasets and they dispose of huge economical (usually public) resources.

Now, we are at a crucial crossroad, we face scientific challenges that need to be addressed from different perspectives by joining multidisciplinary efforts. And thus, the increase in complexity is served. If we want to have an impact and make our data available to the world, we need to aim at having pan-national repositories, in the sense that we want to disseminate the data as much as possible. With the opportunities provided by Artificial Intelligence (AI), several profit and non-profit international initiatives were launched. For data repositories, we can find ZENODO⁴ from CERN, DRYAD⁵ from NCSU, Dataverse⁶ from Harvard University, Figshare⁷ from Digital Science-Macmillan, or Menderley Data⁸ from Elsevier, among others. All of them are multidisciplinary and the differences mainly rely on the type and format of the data, the size of the datasets, and the storage and diffusion licences. These repositories are not merely storages of metadata, they have a rigorous management and stewardship of these valuable digital resources to the benefit of the entire academic community. Although these are great news, several ethical concerns are raised, such as for how long these datasets need to be stored, can data be retrieved, do the platforms meet the requirements of the different Funding Agencies, can they integrate with other platforms, what is their cost, are they accessible, etc. Many universities, research centres and even governmental institutions have adhered to some of those. For instance, in our case, the Generalitat de Catalunya through the CSUC (Consorci de Serveis Universitaris de Catalunya) did so to Dataverse, and this platform is available for use in all our university libraries. This platform supports the implementation of FAIR Data. But there is still a long road to pave in how to improve the data visibility and accessibility to everybody in a user-friendly manner to encourage integrative and exploratory behaviours, while preserving privacy.

In the case of sharing software, it depends very much on the scientific field. There are the basics such as GitHub⁹, which is a website and cloud-based service that helps developers store and manage their code, as well as track and control changes to their code. This is an AI-powered developer platform to build, scale, and deliver secure software, with a user-friendly interface that even neophyte coders can take advantage of it. Code scripts are usually shared as well in labs' web pages,

where people can contact us for assistance. Many other initiatives helping to bridge disciplines are out in the cloud. For instance, the Fiji project¹⁰ funded by the National Institute of Health (NIH, US) is driven by a strong desire to improve the tools available for life sciences to process and analyse data. Fiji is an open-source image processing package based on ImageJ2, and provides many bundled plugins that can be updated with your own one's. It features thousands of plugins that aid in scientific image processing and analysis. It is a beautiful example of a truly international collaboration, since all the new plugins improve the package and can be immediately shared. This has provided biologists with unprecedented advantages to obtain biological insights from large high-resolution imaging datasets when dealing with 3D-images. However, you need to have some computing knowledge and clear criteria of your needs to navigate through all these web sites if you do not want to get lost with so many options.

3. Open-Access publications: do not fool ourselves

Academic publishing has drastically changed over the centuries. Just in the last few years, we have transitioned from paper-printed journals, to online-only journals and to preprints, and witnessed big innovations. All these changes deeply modify the day-to-day life of scientists and how science is made and shared, setting new challenges for science and society. The Open Science approach encourages scientists to publish in Open Access (OA), meaning that your research needs to be immediately available to everyone. But what are Open Access publications? It refers to freely available, digital, online information. Open Access scholarly literature is free of charge and often carries less restrictive copyright and licensing barriers than traditionally published works, for both the users and the authors. This can be done by using i) available repositories where you can upload your non-peer reviewed article that will be widely disseminated, ii) open-access born academic peer-reviewed journals, or iii) regular academic peer-reviewed journals where you pay a plus for having your article in Open Access.

Examples of available repositories in biomedicine are the preprint servers for biology and health sciences launched by Cold Spring Harbor such as bioRxiv¹¹, and medRxiv¹², that got a lot of visibility during the last COVID-19 pandemics. And similar servers are available in many other fields. Although this has encouraged many of us to quickly share our data while the manuscript is under peer-revision in a scientific journal, we need to consider them as work in progress that most probably will be improved upon revision. We can all remember that medRxiv was very much present in all press-media during COVID-19 pandemics, and how some of the results that were published there turned out not to be as good as expected. Science requires time, and although sometimes peer-review is

harsher than what it should be, it has to be recognized that in most cases it makes our work better.

Some academic journals were already born as Open Access, such as the Public Library of Science, PLOS¹³, launched in 2006. It is a non-profit publisher of OA journals in science, technology, and medicine and other scientific literature, under an open-content license. More recently, with the mission of bringing real change in the way the results of research were reviewed and communicated, eLife¹⁴ was created. It was established at the end of 2012 by the Howard Hughes Medical Institute, Max Planck Society, and Wellcome Trust, inspired by research funders and led by scientists. These academic initiatives have been accompanied by the rising of many more OA journals, together with many predator ones, which take advantages of the Open Science movement to make their own profit. PLOS and eLife initiatives, together with the scientific journals associated to academic societies, were successful in driving some of the classical academic journals towards the Open Access mind. Now, some of the highly reputed ones have a sister journal which is entirely Open Access (Scientific Reports for Nature Publishing Group; Cell Reports for Cell Press; etc.). However, the vast majority has not.

One of the perversities that arose during these last years is that scientists are asked to publish their work in good journals and in Open Access, which implies a huge increase in the publication fees but not accompanied with an increase in the funds to pay them. Let me explain myself. Academic journals in general apply the publication fee upon acceptance of the manuscript. In biology, this fee can be up as 6000€/article, and even in PLOS or eLife, this fee is around 2500€/article. As it is well known, to peer-review means that we scientists peer-review for free, because this is a collective communal duty. Thus, we pay for publishing upon acceptance from our competitive funds and our host institutions pay for journal subscriptions. In addition, to have an article in Open Access in the journals that are not, authors can be charged up to 10000€ to make research papers free to read. Thus, the result is that to publish a good quality article in Open Access can double the price of it. This makes unsustainable for many scientists to play in the main league and to have all articles in Open Access, and in my opinion, this cannot be changed by single individuals or single countries. Germany, with all its scientific power and led by Dr Angela Merkel, decided few years ago to face the abuses of the big scientific publishers such Elsevier (Vogel, 2014; Schiermeier and Mega, 2017). For this, they developed the Projekt DEAL¹⁵ that is a project led by a consortium of German research institutions to negotiate nationwide licenses with the largest academic publishers such as Elsevier, Wiley and Springer Nature, using the boycott to press them to agree to fair publishing charges. This pressure allowed DEAL to get some agreements such as the Publish and Read Fee (PAR Fee) which rolls up the costs for both open access publishing services and comprehensive reading access into one fixed,

all-inclusive fee paid for each article published by a German (corresponding) author. But this strategy can succeed only when your specific weight as a customer is very big, for instance the size of German science. In the Netherlands, 14 universities directly negotiated an Open Access agreement with Elsevier (a Dutch enterprise) and did not get the same deal; they got the compromise of 30% of Dutch papers in open access by 2018 (Elsevier, 2023).

Why do we still want to publish in such journals even though we have to pay them large amounts of money to publish in Open Access? Why do so if we all know that the outputs from scientific research are many and varied, including: research articles reporting new knowledge, data, reagents, and software; intellectual property; and highly trained young scientists? It is undoubtedly problematic to take a journal metric or brand as a proxy for the impact or quality of an individual article. Nevertheless, especially in the formal and natural sciences, publications in prestigious and highly cited journals are still considered to be an essential factor in evaluations, for the acquisition of external funds, and to be considered for science awards or for professorships. In 2013, the principles of The San Francisco Declaration on Research Assessment, DORA¹⁶, were established, with the mission of advancing in practical and robust approaches to research assessment globally. The main goal was to fight the leading role of the Journal Impact Factor for performance evaluations. Funding agencies, universities and research centres, and scientists, we all have the desire, and need, to assess the quality and impact of scientific outputs, and therefore to measure it accurately and evaluated wisely. However, the DORA principles are not yet universally implemented at the institutional and funder level. We need to accept that although easy to say this is not such an easy task; how to give proper credit to people for the research they accomplished? We like it or not, all funding agencies and the job market still work with journal metrics or brand parameters, and by job market, I mean academia as well. We train people in our research groups, PhD students and postdocs that will be out looking for a job and will be evaluated with a broader scope, but still, the quality of the work will be mainly assessed by the papers they published. We cannot do experiments with them, and most importantly, we do not want! When you got tenure, it is easy to say that we will boycott some high-standard journals because we do not like their gaming, but it is more difficult to defend this position when you know the importance of publishing in someone else's cv, and how this may affect to your funding resources.

4. What's next?

The different administrations are pushing scientists in the hope of having a full Open Science circuit in a near future. We have the guidelines from the European Commission, a new Spanish Law for Universities and the Catalan Law for Research

highly encouraging the use of Open Access publications as an indicator of the scientific performance. In the University Pompeu Fabra, we provide the needed help for undergoing this scientific transformation¹⁷. We have not only guidelines but help from the library personnel, who are extremely well trained in this matter. The scientific community is moving in an unprecedented manner towards Open Science, sharing as best as we can all the resources. Most of us strongly believe that big discoveries and breakthroughs in science require collaborative efforts and we are happy to be part of it. However, it is difficult to do as much with so little. And specially, we need to stop to be hypocritical in how the system can change. DORA was created in 2013 with the support of a huge number of scientists, universities and research institutions, and still 10 years later funding agencies rank us according to the number of papers in high-impact factor journals and how many big grants we previously got as top criteria. I think that if we want Open Science to succeed, we all need to be more self-aware with the things we say. And who runs the politics and writes the guidelines should know the economical and personal implications. If not, we will miss the opportunity to transform, that in my opinion, it is what really matters.

Endnotes

- 1 Ensembl: <https://www.ensembl.org>
- 2 ENCODE: <https://www.encodeproject.org/>
- 3 EOL, Smithsonian: <https://eol.org/>
- 4 ZENODO from CERN: <https://zenodo.org/>
- 5 DRYAD from NCSU: <https://datadryad.org/>
- 6 Dataverse from Harvard University: <https://dataverse.harvard.edu/>
- 7 Figshare from Digital Science-Macmillan: <https://figshare.com/>
- 8 Mendeley Data from Elsevier: <https://data.mendeley.com/>
- 9 GitHub: <https://github.com/>
- 10 Fiji project, funded by the National Institute of Health, NIH, US: <https://fiji.sc/>
- 11 bioRxiv, by Cold Spring Harbor: <https://www.biorxiv.org>
- 12 medRxiv, by Cold Spring Harbor: <https://www.medrxiv.org/>
- 13 Public Library of Science, PLOS: <https://plos.org/>
- 14 eLife: <https://elifesciences.org>
- 15 Deal project: <https://deal-konsortium.de/en/about-deal>
- 16 DORA: <https://sfdora.org/>
- 17 Pompeu Fabra University Open Science Webpage: <https://www.upf.edu/en/web/biblioteca-informatica/temes/ciencia-oberta>

References

- Consortium, I. H. G. S., Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. <https://doi.org/10.1038/35057062>
- Elsevier (2023). Agreement between The Netherlands and Elsevier. *Elsevier*. <https://www.elsevier.com/open-access/agreements/vsnu-nl>
- Jacobsen, A., Azevedo, R. de M., Juty, N., Batista, D., Coles, S., Cornet,

R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., et al. (2020). FAIR Principles: Interpretations and Implementation Considerations. *Data Intelligence* 2(1-2), 10–29. https://doi.org/10.1162/dint_r_00024

Schiermeier, Q. and Mega, E. R. (2017). Scientists in Germany, Peru and Taiwan to lose access to Elsevier journals. *Nature* 541, 13–13. <https://doi.org/10.1038/nature.2016.21223>

Vogel, G. (March 27, 2014). German University Tells Elsevier 'No Deal'. *Science.org*. <https://www.science.org/content/article/german-university-tells-elsevier-no-deal>

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Santos, L. B. da S., Bourne, P. E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

CV

Cristina Pujades

- cristina.pujades@upf.edu
- <https://orcid.org/0000-0001-6423-7451>
- Full Professor at the Universitat Pompeu Fabra and ICREA Academia awardee from the Generalitat de Catalunya. After my PhD in Biology from the Universitat de Barcelona, I did a postdoctoral stay at the DFCI-Harvard Medical School (Boston). In 1995, I moved to Paris as a postdoc at the Ecole Normale Supérieure, and in 1999 I got a permanent position at Sorbonne University. From that time, I have been fascinated by developmental neurobiology, and how to move from DNA and genes to the generation of form. This has been the main focus of my latest scientific career. In 2002, I had the opportunity of joining the Department of Medicine and Life Sciences at the Universitat Pompeu Fabra. In my lab, we are interested in understanding how spatiotemporally coordinated cell progenitor specification and differentiation occur during morphogenesis to construct a functional brain (<https://pujadeslab.upf.edu/>). Currently, I am serving as Vice-Rector for Research at the Universitat Pompeu Fabra.