

# Open Science: a tutorial for the database systems community

Emma Lazzeri<sup>1</sup>[0000-0003-0506-046X] and Paolo Manghi<sup>1</sup>[0000-0001-7291-3210]

<sup>1</sup>Institute of Information Science and Technologies (ISTI), National Research Council of Italy (CNR),  
Via G. Moruzzi, 1, 56124, Pisa, Italy.  
emma.lazzeri@isti.cnr.it

**Abstract.** This Tutorial, presented at the 28th Symposium on Advanced Database Systems (SEBD2020), aims at introducing the motivations and main features of Open Science, linking it to the research integrity and reproducibility of science, with a focus on the challenges in the ICT and database systems domains.

**Keywords:** Open Science, Open Access, Research Data Management, FAIR principles.

## 1 Motivations

The way research is conducted is currently influenced by external yet connected factors: scientific journals market and evaluation criteria. Research communication system today mainly relies on business models that allow researchers to read articles only if their institution pays subscriptions to gain access to a limited set of scientific journals. The global business of scientific communication is estimated to be as worth as US \$ 10 billion per year [1], with an increasing trend.

This system is strongly linked to the current research evaluation models, mostly relying on bibliometric indexes based on citation metrics that present several limits, such as the Journal Impact Factor and the H-index [2,3]. Besides the intrinsic drawbacks in the use of these indexes to assess a researcher, this system limits the mechanism of verification and control of related results by "peers" and the whole scientific community, as well as the fertilization of new ideas by obstaculating the access to scientific papers to the broader audience. Furthermore, we currently neglect to give access to fundamental parts for the proof of what is reported in the published articles, simply because they are not involved in the research assessment: research data, software, methodologies, and other results. It is also worth noting that researchers currently act as editors, reviewers, and authors of scientific papers, without being remunerated by publishers that make profit also based on the unpaid raw material scientists produce [4].

## 2 The Open Alternative

There is indeed an alternative way of doing research: Open Science. Open Science is based on transparency and collaboration and aims at overcoming the barriers to the research results sharing and facilitating the dissemination of knowledge. Open Science means opening each step of the research life cycle. It means sharing research results as much as possible. Transparency, reproducibility, collaboration, inclusiveness, accessibility, accuracy, and re-use are the key principles of Open Science that steps from the concept that the research that is funded with public money has to be made immediately available to the community: every EU citizen has the right to access and benefit from knowledge produced using public funds [5].

The European Commission and a long list of International Funders made a clear choice towards Open Science, as it means enabling broader access to publicly funded research results and therefore helps to build on previous research results, encourage collaboration and avoid duplication of effort, speed up innovation, and involve citizens and society. Open science is defined as an “umbrella term”, comprising different elements: from open access to research results (literature, data, software, etc.), to open peer review, open methodologies, protocols, workflows, from open education to citizen science. These elements need to be embedded in a system where research infrastructures and a new evaluation model go hands in hands with research integrity.

One important aspect of embedding open science in the everyday life of researchers is research data management, which implies a structured way of completing the research data lifecycle with the main objective of delivering re-usable research data that can be shared with others. A good research data management needs to follow the FAIR principles, a set of good practices to help making data Findable, Accessible, Interoperable and Reusable [6].

Openness and FAIRness are therefore the means to make science more transparent and reproducible, repeatable, replicable, reusable. In this view, research data is just one of the resources involved. Open Science is about each element of research: data, software, publications, services, etc. There is a general need to identify Open Science resources and how these are related, to ensure their openness and FAIRness. In this context, the definition and development of standards, tools and research infrastructure eliminating the barriers and facilitating the work of scientists by embedding open science good practices in their daily work is key. Several tools and infrastructures are already in place, other needs to be developed.

In this context, one of the latest initiatives of the European Commission is the launch of the European Open Science Cloud, that aims at creating a virtual research environment to access and interoperate research data and other research outputs in Europe across the different disciplines [7-9].

### **3 Opportunities and challenges for the database systems community**

Open Science in the Database Systems sector deals with sharing software in a way that makes it reproducible, preservable, and citable, but also with new and challenging research opportunities linked to establishing infrastructures that can support best practices for research transparency and collaboration.

The “R\* of Science” deal with actions that should be the basis of scientific method [10,11] . Repeating science deals with defending the thesis (repeat the same experiment with the same setup in the same lab). The method researchers claim should also be Replicable in order to be certified by others (same experiment and setup, independent lab). Reproducibility of science introduces variations in some of the aspects of research methods (experiment setup or lab). Finally, Reusing research results deals with the transfer of knowledge to enable different experiments, also by others.

Best practices in the ICT domains already exist, ranging from software collaborative development and publication, software and data papers drafting, preprint and postprint selfarchiving, dataset FAIR management and sharing and interlinking of results.

However, work still needs to be done on reproducibility. Research in this sector includes challenging aspects as the definitions of standards and templates for reporting methods, provenance and tracking, the workflow/script automation, design and development of tools and platforms for capturing, tracking, structuring, organising assets throughout the whole project research cycle.

The tutorial presented at SEBD2020 by the authors is available in open access in Zenodo [12].

#### **References**

1. Schimmer, R., Geschuhn, K. K., & Vogler, A. (2015). Disrupting the subscription journals' business model for the necessary large-scale transformation to open access. <https://doi.org/10.17617/1.3>
2. Okubo, Y. (1997), "Bibliometric Indicators and Analysis of Research Systems: Methods and Examples", OECD Science, Technology and Industry Working Papers, No. 1997/01, OECD Publishing, Paris, <https://doi.org/10.1787/208277770603>
3. Haustein S., Larivière V. (2015) The Use of Bibliometrics for Assessing Research: Possibilities, Limitations and Adverse Effects. In: Welpé I., Wollersheim J., Ringelhan S., Osterloh M. (eds) Incentives and Performance. Springer, Cham. [https://doi.org/10.1007/978-3-319-09785-5\\_8](https://doi.org/10.1007/978-3-319-09785-5_8)
4. Stephen Buranyi, "Is the staggeringly profitable business of scientific publishing bad for science?", The Guardian, June 27, 2017, <https://www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science>
5. Neelie Kroes, Vice-President of the European Commission responsible for the Digital Agenda The Challenge of Open Access Launch of OpenAIRE, the European infrastructure for open access publishing of research results Ghent, 2 December 2010, [https://ec.europa.eu/commission/presscorner/detail/en/SPEECH\\_10\\_716](https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_10_716)
6. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
7. European Commission - EOSC, <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>
8. EOSC Secretariat, [www.eoscsecretariat.eu](http://www.eoscsecretariat.eu)
9. EOSC portal, <https://www.eosc-portal.eu/>
10. Jill P. Mesirov, Accessible Reproducible Research, Science, 22 JAN 2010 : 415-416, <https://doi.org/10.1126/science.1179653>
11. Benureau Fabien C. Y., Rougier Nicolas P., Re-run, Repeat, Reproduce, Reuse, Replicate: Transforming Code into Scientific Contributions, Frontiers in Neuroinformatics, v. 11, 2018, p.69, <https://doi.org/10.3389/fninf.2017.00069>
12. Link to the tutorial presentation, <https://doi.org/10.5281/zenodo.3904168>