Munin Conference on Scholarly Publishing

# Open Science Infrastructure as a key component of Open Science

*Jadranka Stojanovski (University of Zadar / Institute Ruđer Bošković, Croatia)*

**Abstract**

The Open Science movement is a response to the accumulated problems in scholarly communication, like the „reproducibility crisis“, „serials crisis“, and „peer review crisis“. The European Commission defines priorities of Open Science as Findable, Accessible, Interoperable and Reproducible (FAIR) data, infrastructure and services in the European Open Science Cloud (EOSC), Next generation metrics, altmetrics and rewards, the future of scientific communication, research integrity and reproducibility, education and skills and citizen science. Open Science Infrastructure is also one of four key components of Open Science defined by UNESCO.

Mainly represented among Open Science Infrastructures are institutional and thematic repositories for publications, research data, software and code. Furthermore, the Open Science Infrastructure services range may include discovery, mining, publishing, the peer review process, archiving and preservation, social networking tools, training, high-performance computing, and tools for processing and analysis. Successful Open Science Infrastructure should be based on community values and responsive to needed changes. Preferably the Open Science Infrastructure should be distributed, enabling machine-actionable tools and services, supporting reusability and reproducibility, quality FAIR data, interoperability, sustainability, long-term preservation and funding.

**Keywords**

open science infrastructure, scholarly communication, open science, EOSC

## Open Science as a response to the scholarly communication crisis

Science is the process of acquiring new knowledge about the world surrounding us, assuming that this world follows certain rules and principles. However, the scientific methods used to discover these rules and principles involve a level of uncertainty. Therefore, the ability to repeat research and compare results is crucial for science, and the present replicability and reproducibility crisis endangers the core values of science: the possibility to assess research findings critically and to prevent the waste of time and resources. The ultimate purpose of science is, therefore, to offer researchers all the data they require to repeat (same team, same experimental setup), replicate (different team, same experimental setup), reproduce (different team, different experimental setup) or reuse previous research results (Manola, 2017; Plesser, 2018) to confirm scientific claims made by other scientists or to advance their research. Although it is tough to measure the present share of irreproducibility in science, especially in view of different definitions and

Stojanovski, J. “Open Science Infrastructure as a key component of Open Science”, presented at the 17th Munin Conference on Scholarly Publishing 2022. Published in *Septentrio Conference Series* 1 (2022). https://doi.org/10.7557/5.6777

https://doi.org/10.7557/5.6777

interpretations of terms, a few studies show worrying results. For example, Baker's (2018) study shows that more than 70% of researchers, mainly from STM disciplines, have failed to reproduce another scientist's experiments, and more than 50% have failed to reproduce their own experiments.

The share of 3R (reproducibility, replicability and repeatability) in scholarship depends significantly on the discipline's characteristics and the study type. Considering the different degrees of standardisation of methods and materials, control over environmental variability and the reliability of statistical methods, Leonelly (2018) calls for the need to look more closely at exploratory and non-standard research conducted on rare, unique, perishable and inaccessible materials in highly variable environmental conditions, where direct reproducibility cannot function as a measure of research quality. However, there are also strong arguments that replicability in the humanities is possible and desirable, especially concerning empirical studies (Peels, 2019), so the discussions about 3R in the humanities and related disciplines are still ongoing.

In order to enable 3R in scientific research, and thus the faster advancement of scholarship and trust, it is necessary to develop a scientific culture that promotes rigorous research design and research practices that improve 3R adoption through education and training. Additionally, the description of methods should be more exhaustive and/or standardised, pre-registration of research should be promoted, research data should be stored in open repositories, and appropriate incentives that value research rigour and the production of reproducible results should be in place. Journals can mandate 3R culture through their editorial policies and the publication of "negative results".

Furthermore, the scientific community continues to face the accumulated problems of the "serials crisis" related to the vast commercialisation of journal publishing, significant price increases, copyright transfer and restrictive licensing. Additionally, the scholarly communication system is still dominated by traditional publishing, mostly in journals and books, which brings a functionality crisis characterised by missing digital modernisation and "researcher time wasted on antiquated procedures, e.g. in discovery, submission or review" (Brembs et al., 2021).

Publishing in peer-reviewed journals benefits scientists for their professional advancement and improves the image of academic institutions and disciplines, as peer review is considered to confirm the quality of reporting on research results. Still, various studies show that the peer review process is flawed, slow, expensive, ineffective, biased, opaque, easily abused, unreliable and without incentives. However, its shortcomings and weaknesses do not reduce its value and relevance. Instead, the "peer review crisis" call for changes and more innovative approaches (Stojanovski et al., 2021).

As a response to all challenges, an Open Science (OS) initiative started in Europe in 2014, seeking to make the research process more transparent, all research outcomes more accessible, knowledge freely available to everybody. The initiative also had among its objectives scientists working more collaboratively and open research infrastructure that will facilitate openness and collaboration at all levels.

According to the 2020–2021 EUA survey of European universities, OS' strategic importance is highly rated but less implemented. More than half of the universities have an open science policy, and the majority of institutions have their own or shared infrastructure supporting Open Access (OA) to research publications (Morais et al., 2021). However, according to the same study, the main barriers to the institutional transition to OS are the absence of incentives to promote OS activities, concerns over the legal framework, and concerns over increased costs.

OS is not so innovative when it comes to conducting science itself ("open science is just science done right", Imming & Tennant, 2018), influenced by technological achievements, as when it comes to how scientists collaborate and share their results with other scientists and society. After fifteen years of supporting OA, the European Commission (EC) has more recently been strongly promoting the principles of OS, advocating the transition to an open and interdisciplinary system of knowledge creation accessible to all stakeholders in society. In summary, the EC defines its priorities to advance the OS agenda as follows:

1. FAIR (Findable, Accessible, Interoperable and Reusable) data;
2. European Open Science Cloud (EOSC) – providing tools and services for research purposes;
3. Next-generation metrics, altmetrics and rewards – development of new indicators and alternative metrics that promote the practice of OS by researchers;
4. The future of scientific communication – principles that will shape the vision of scientific communication in the next 10–15 years;
5. Research integrity and reproducibility of scientific results;
6. Education and skills – the acquisition and practice of OS skills as an integral part of the researcher's professional training and development,
7. Citizen science.

The Conclusions of the EU Council (2022) call for reforming the research assessment system, which is currently focused on using inappropriate quantitative indicators and journal publishing. Given that such an approach respects only a narrow range of research outcomes, giving a distorted picture of the quality, reproducibility and integrity of research, it is recommended to use an assessment system that will include other results and processes of scientific research and promote an early exchange of knowledge and cooperation.

## Open Science Infrastructure (OSI)

The availability of dependable, secure, and robust infrastructure and value-added services integrated into every stage of the research process is essential for encouraging researchers to adopt open science practices (Manola, 2017). In their 2020 Report, SPARC Europe defines open access and open science/scholarship infrastructure (OSI) as sets of services, protocols, standards and software contributing to the research lifecycle, from collaboration and experimentation through data collection and storage, data organisation, data analysis and computation, authorship, submission, review and annotation, copyediting, publishing, archiving, citation, discovery and more (Ficarra et al., 2020).

According to UNESCO Recommendations (2021), OSI is one of the four key components of open science. Shared research infrastructures (virtual or physical) could be represented by major scientific equipment or sets of instruments, OA journals and publishing platforms, repositories, archives, current research information systems, open computational and data manipulation services in open labs, and virtual research environments.

Mainly represented among OSIs are institutional or thematic repositories for publications (including preprints), research data and source codes, and software. Among the crucial elements of OSI are persistent identifiers, allowing identifying unambiguously scientific

https://doi.org/10.7557/5.6777

objects. In addition, OSI provides essential open and standardised services to manage and provide access, portability, analysis, and federation of data, scientific literature, thematic science priorities or community engagement. While various repositories are tailored to the particulars of the objects they hold (publications, data, or code), to local conditions, user needs, and the demands of research communities, they should all adhere to interoperable standards and best practices to guarantee that the content in repositories is adequately vetted, discoverable, and reusable by both humans and machines. Furthermore, OSI is frequently the outcome of community-building initiatives, which are essential for their long-term viability. As a result, they should be non-profit, community-owned and, to the greatest extent possible, offer perpetual and unlimited access to all members of the public (UNESCO, 2021).

The range of OSI services is broad and includes discovery (search and retrieval), mining, publishing (journals, books, preprints), the peer review process, digital object archival and preservation (permanent links), social networking tools, training, high-performance computing, and processing and analysis tools.

Successful OSI should be based on community values and responsive to changing needs. According to the Principles of Open Scholarly Infrastructure, stakeholder-driven infrastructure and transparent procedures build more confidence. The community should also drive the regulations and policies relevant to the infrastructure, including non-discriminatory membership (Bilder & Neylon, 2015). For infrastructures to be successful, the policies that support them must also work together. The emphasis is on interoperability and the accompanying guidelines, which are released either at the data/metadata level or the service level due to diversity and various adoptions of solutions and requirements (Manola, 2017).

To accelerate and support the current transition to more effective OS and open innovation, the European Open Science Cloud (EOSC) was officially launched in 2018. It should enable trusted access to sustainable services, systems and the reuse of shared research data across disciplinary, social and geographical borders and support sustainable infrastructures and core resources. EOSC could be defined as a "researchers-centred, value-based, excellence as well as impact-driven area, in which researchers, knowledge and technology are supported and can circulate freely." Expected EOSC outcomes are the deployment of OS principles and identification of best practices, deployment of the core components and services of EOSC and federation of existing data infrastructures in Europe, working towards the interoperability of research data according to the FAIR principles, and establishing a monitoring mechanism to collect data and benchmark investments, policies, digital research outputs, OS skills and infrastructure capacities related to EOSC (Barker, 2021).

Preferably the infrastructure should be distributed, instead of centralised, supporting local priorities and languages. In addition, enabling machine-actionable tools and services, supporting reusability and reproducibility, correct, accurate and reliable FAIR data, interoperability, sustainability, long-term preservation, and appropriate funding are desirable features of OSI.

## How open is Open Science?

There is no doubt that researchers strongly support OS principles. Gaining more reliability, trust and equality by enabling openness and transparency is the goal everyone would support. Still, providing OA to all publications, including educational materials, research data, software, code, and methods, requires time, resources, user-friendly infrastructure, training and support. Additionally, sometimes best intentions could be questionable, risky, not realistic or misimplemented when it comes to the specific subdiscipline or research community.

### *Funding and Economic Inequalities*

Economic inequalities between countries are significant, and existing funding models through research projects and grants deepen these differences. OS initiatives and infrastructures in some countries rest on the shoulders of a small number of individuals and their exclusively voluntary work. Such countries are underrepresented in reporting, monitoring or evaluating the level of OS adoption, despite many successful bottom-up OS services, initiatives and resources.

This is shown by numerous examples of diamond national or regional journals that, despite quality editorial policies, struggle to compete with the mainstream journals of larger publishers. Moreover, in a situation where, for example, the editorial work is primarily voluntary, and journal subsidies are insufficient or non-existent, such journals can hardly adopt additional technical requirements.

Also, the power imbalance is visible when sharing other research outcomes, such as software, codes or datasets, which require significant time investments and additional training. In a situation where scientists are fighting for their jobs and funding within existing systems that mainly value publications, it is particularly difficult to motivate them to share other research results.

In addition, the criteria for reporting on the degree of adoption of OS principles are often binary and designed according to wealthier countries/institutions, while in countries with a lower economic status, there are many different layers of OS adoption. Existing OSIs should be less discriminatory towards journals, infrastructures and services from less developed countries, offering adequate support.

### *Ownership*

Although the trends of privatisation and commercialisation are difficult to stop, OSI should be based on open source systems and focused on the interests of the broader scientific community. It is also essential to maintain community ownership of infrastructure and facilities.

### *Publication costs*

Although most scientists support the principles of open science, decades of "closed" science have woven certain patterns into research practices, and it will take more time to change them. For example, the ubiquitous concept of prestigious journals in which it is necessary to publish if a scientist considers himself successful has led to even greater commercialisation in scientific publishing. In Europe, big deals have been replaced by different types of transformative agreements that underpin gold and hybrid journals from major publishers, including limited or unlimited OA publishing rights, APC discounts, and institutional repository archival rights (with or without embargo).

This way, the "journal crisis" has been replaced by the "affordability crisis". Although today most of the papers published in journals are in OA, the price the research community pays

https://doi.org/10.7557/5.6777

is even higher, and the fact that subscription prices are "replaced" with Article Processing Charges (APC) by many publishers is making the financial burden even heavier and inequities even bigger (Ross-Hellauer, 2022). In order to publish in open access, scientists must have the funds available to ensure the payment of high APCs. In the opposite case, as the example from Serbia shows, scientists are forced to pay the costs of publishing in open access from their own pockets, further putting them in an unequal position (Sevkušić, 2020). Otherwise, to advance in their academic career, scientists are forced to publish in prestigious journals in closed access, making their work less visible, read and cited.

In low-income countries, which cannot afford more expensive transformative contracts, scientists must pay APC or publish in diamond journals, which most existing assessment systems consider an inferior publication channel. Early career researchers who often change workplaces and do not have financial resources face particular difficulties. Even at institutions or grants with dedicated funds for OA publishing, prevalent journal business models encourage unhealthy competition among authors.

Although open science advocates claim that there is always the possibility of publishing a preprint or depositing a certain version of a work in an institutional repository (the so-called green path of OA) without additional fees, practices show that research assessment bodies do not adequately value such "alternative" approaches. Additionally, not every institution has a well-managed repository, and the acceptance of preprints is unequally represented in different scientific fields.

### *Author's rights*

In some circumstances, the Rights Retention Strategy, mandating author accepted manuscript depositing with Creative Commons Attribution (CC BY), could be challenging to negotiate with the publisher and expensive to pay. Therefore, legal support should be provided to the author(s).

### *Data*

Although existing research does not confirm a significant possibility of scooping when it comes to preprints or datasets sharing, a certain risk that someone could publish it or monetises it in some other way still exists, especially when it comes to scientific communities with fewer resources and without adequate legal and technological support.

Additionally, the possibility of leaked sensitive information exists despite the societal advantages of open data exchange. The tools and procedures used to deidentify data are still in their development or are not equally suitable for all disciplines and all kinds of data, and the researchers usually lack the necessary training to handle data safely. Identification and deanonymisation of the data may be possible also after linking them through several datasets (Khalil, 2022).

### *Peer review*

There are worries about open review processes, which may help foster more unethical behaviour. This can certainly be present to a greater extent in small and non-English research communities, where there are very few experts within a particular scientific discipline and where it could be challenging to conduct an independent and objective evaluation. Also, early-career researchers could be restrained in criticising the works of senior scientists, who decide on their employment, project approval and career. A possible solution lies in the optional disclosure of the reviewer's identity so that the open review process includes only the transparency of the peer review process and the publicly available reports of the reviewers. With time, academic and societal culture could change and develop in a direction which will enable complete transparency.

### *Research assessment*

The situation in low-income countries is further aggravated by the higher acceptance of the traditional research assessment systems, primarily based on the number of papers in prestigious journals. Therefore, the proposed reform of the research assessment system by the EU Council should be implemented carefully, consistently and globally.

## Conclusions

Although OS can solve some of the accumulated problems of scientific communication, it certainly cannot solve all the problems deeply embedded in the culture of scientists and society. Moreover, in practice, OS can reinforce existing biases and inequalities in the scientific community.

Therefore, further work to advance fundamental principles like diversity, inclusivity, gender equality, and ethics, citizen participation in the scientific process, national OS policies, a combination of top-down initiatives on open science from policymakers and bottom-up initiatives from the community, the development of sustainable open infrastructures, and changing of research culture is needed. All steps scientists take to make science more open must be valued, continuing to remove the economic, societal, structural and technical barriers that currently slow or prevent the adoption of OS.

**References**

Barker, M., Manola, N., Gaillard, V., Kuchma, I., Lazzeri, E., Stoy, L., & Piera, J. (2021). Digital skills for FAIR and open science: Report from the EOSC Executive Board Skills and Training Working Group. European Commission. DOI: https://doi.org/10.2777/59065

Baker, M. 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454 (2016). https://doi.org/10.1038/533452a

Bilder G, Lin J, Neylon C (2015) Principles for Open Scholarly Infrastructure. DOI: http://dx.doi.org/10.6084/m9.figshare.1314859

Brembs, B., Huneman, P., Schönbrodt, F., Nilsonne, G., Susi, T., Siems, R., Perakakis, P., Trachana, V., Ma, L. & Rodriguez-Cuadrado, S. (2021). Replacing academic journals. Zenodo. DOI: https://zenodo.org/badge/DOI/10.5281/zenodo.5793611.svg

Ficarra, V., Fosci, M., Chiarelli, A., Kramer, B., & Proudman, V.. (2020). Scoping the Open Science Infrastructure Landscape in Europe. Zenodo. DOI: https://doi.org/10.5281/zenodo.4159838

Melanie Imming, & Jon Tennant. (2018). Sticker open science: just science done right (ENG). Zenodo. https://doi.org/10.5281/zenodo.1285575

Khalil AT, Shinwari ZK and Islam A (2022) Fostering openness in open science: An ethical discussion of risks and benefits. Front. Polit. Sci. 4:930574. DOI: https://doi.org/10.3389/fpos.2022.930574

https://doi.org/10.7557/5.6777

Leonelli, Sabina (2018) Re-Thinking Reproducibility as a Criterion for Research Quality. [Preprint]: http://philsci-archive.pitt.edu/14352/1/Reproducibility_2018_SL.pdf

Manola, N. (2017). e-infrastructures: The starting blocks for open science and innovation the OpenAIRE case. Paper presented at the Digital Presentation and Preservation of Cultural and Scientific Heritage, 6 19-27.

Morais, R., Saenen, B., Garbuglia, F., Berghmans, S. and Gaillard, V. (2021). From principles to practices: Open Science at Europe's universities 2020-2021 EUA Open Science Survey results. European University Association. DOI: http://doi.org/10.5281/zenodo.5062982

Peels, R. Replicability and replication in the humanities. Res Integr Peer Rev 4, 2 (2019). DOI: https://doi.org/10.1186/s41073-018-0060-4

Plesser H.E. (2018) Reproducibility vs. Replicability: A Brief History of a Confused Terminology. Front. Neuroinform. 11:76. DOI: https://doi.org/10.3389/fninf.2017.00076

Ross-Hellauer T., Reichmann S., Cole N.L., Fessl A., Klebel T., Pontika N. (2022). Dynamics of cumulative advantage and threats to equity in open science: a scoping review. R. Soc. Open Sci. 9: 211032. DOI: https://doi.org/10.1098/rsos.211032

Sevkušić, M. (2020). Helping researchers handling APCs in Serbia. EIFL Blog. https://www.eifl.net/blogs/helping-researchers-handle-apcs-serbia

UNESCO (2021). UNESCO Sets Ambitious International Standards for Open Science. UNESCO press release 2021-128, November 25, 2021. Available online at: https://www.unesco.org/en/articles/unesco-sets-ambitious-international-standards-open-science