Recent findings and policy messages for open science

Open science is a means and not an ends. Open science strategies and policies are a means to support better quality science, increased collaboration, and engagement between research and society that can lead to higher social and economic impacts of public research.

Open science is more than open access to publications or data; it includes many aspects and stages of research processes. It is important to remember that open science is a broader concept that also includes the interoperability of scientific infrastructure, open and shared research methodologies (such as open applications and informatics code), and machine-friendly tools allowing, for example, text and data mining.

Policies to promote open data are less mature than those to promote open access to scientific publications. While the principle of open access to scientific data is well established in OECD countries, the scope of access varies greatly across countries. This is due to the fact that data sets are not as easily identified and defined as scholarly research articles. Diversity of scientific data and differing traditions and standards in their treatment are also issues. Some of the additional challenges related to data sets include the definition of ownership of large-scale data sets, potentially collected by machines or software providers; privacy; confidentiality; and even national security issues. In addition, certain classes of data, such as medical records, are particularly sensitive due to privacy issues.

***Open science policies should be principle-based but adapted to local realities*.** Open science policies require a diversity of approaches, taking into account the needs of the different actors involved in research projects. For example, if a research project involves business sector partners and commercial interests are present, the requirements for sharing research results may be different from the case in which only public actors are involved. In other cases, privacy or confidentiality concerns may apply to the treatment of certain classes of individual data.

Better incentive mechanisms to promote data-sharing practices among researchers are needed. While all public sector researchers have an interest in maximising the sharing of published research articles, the same is not true for research data sets, especially at the pre-publication stage. In addition, data cleaning and curation (for example, by developing metadata) is a time-consuming activity that is rarely acknowledged in evaluation mechanisms or grant allocation procedures. Most evaluations of universities and researchers are almost entirely based on teaching and bibliometric indicators, attributing little value to the sharing of pre-publication inputs and post-publication outcomes, such as data and other relevant information. Extending citation mechanisms to data sets can partly address this issue.

Data-related skill development is essential. Researchers' skills needed for sharing articles or data sets openly on line are unevenly distributed. Some disciplines such as computer science or physics may have a longer tradition of uploading research material on repositories and curating and maintaining large data sets. Researchers in other disciplines, however, may need to be trained to develop the necessary skills to make open science happen. At the same time, students and citizens need to acquire the skills to take advantage of, use and reuse data set shared by the research community. Some countries are currently developing data science curricula to address this issue.

Training of and awareness-raising among researchers is important for the development of an open science culture. Recent surveys on the behaviour of scientists reveal that not all researchers are necessarily aware of the possibilities offered by open science. In some countries, different institutions regularly organise workshops and training sessions to make researchers aware of these possibilities.

***Repositories and online platforms will not have impact if the information they contain is not of good quality*.** If repositories are not user-friendly and the data sets they contain have not been properly cleaned and curated, or the metadata have not been sufficiently developed, it may be

difficult to maximise their usage.

The long-term preservation costs of openly available research output need to be considered. Open access is not without costs. Many governments and research institutions are currently bearing the costs of offering open access to articles and to data, as well as the costs of storage and the preservation of data sets on line. Given the rapidly increasing amounts of data, public institutions will be challenged to find sustainable funding and business models. Public-private partnerships with private service providers may offer innovative solutions.

Clear legal frameworks for the sharing of publications and reuse of data sets are needed at the national and international levels. A lack of clarity on the interpretation of national and international legal frameworks may prevent the sharing or reuse of research results. In addition, clear guidelines around text and data mining are needed, as this tool will become increasingly used by researchers in the future. Some OECD countries are currently discussing or have recently modified national legal frameworks to make them increasingly open science-friendly.

Consultative approaches that involve all relevant actors for open science are a key component of successful open science strategies. Open science efforts involve different communities and different actors: researchers, governmental institutions, universities and research centres, libraries and data centres, private non-profit organisations, business sector organisations including private academic publishers, supra-national entities, citizens, etc. These actors do not necessarily have the same incentives, goals or expectations. A successful strategy needs to take into account this diversity, and react accordingly.

International collaboration in the area of open science is necessary to address global challenges. International collaboration is becoming more important than ever, as publications and data in electronic form travel across national frontiers. Shared and interoperable infrastructure is necessary to disseminate research results and promote scientific collaboration. Such efforts can help avoid the duplication of effort, as well as helping share the risks or the associated investments. In addition, BRIC countries Brazil, China and India are also adopting open science policies and data infrastructure roadmaps. International co-ordination and co-operation in this area will become even more important as the global production of knowledge and R&D increasingly shifts towards the emerging economies. Furthermore, tackling global challenges will require greater access to and sharing of national public research data sets – and consequently, greater co-operation at a global level.

Policy makers need to promote openness in science while at the same time preserving competition. Competition is a key aspect of the scientific enterprise: pushing for open access and open data too early may be counterproductive in some cases.