FAIR, FRAND and open - The institutionalization of research data sharing under the EU data strategy

van Eechoud, M.

DOI
10.4337/9781035310869.00047

Publication date
2023

Document Version
Final published version

Published in
Improving Intellectual Property

31. FAIR, FRAND and open – The institutionalization of research data sharing under the EU data strategy

Mireille van Eechoud

1. INTRODUCTION

Long gone are the days when the exchange of research data took place almost exclusively via informal networks of academics. In Europe, access and use of publicly funded research data is subject to institutionalization and regulation by the European Union (EU), despite its modest legislative competences in science. Furthermore, the terms and conditions under which academics must make research data available are increasingly informed by, or even set by, European institutions. The roll-out of the so-called European Strategy for Data plays a large role in this: in particular, the Open Data Directive and the soon to be adopted Data Governance Act and the Data Act. What key concepts feature in these instruments, and where does this leave researchers and their academic freedom?

2. NOTHING NEW UNDER THE SUN?

Institutional arrangements for data sharing are not new. A good example is CERN, the international joint research centre focused on high-energy particles and famous for its Large Hadron Collider. CERN has always been a site for collaborative work and arrangements for the sharing of data and research outputs. More recently, CERN instituted 'open data' policies making part of the data generated freely available to the world. From its inception, CERN was built on the idea of exchange. Under the 1953 Convention for the Establishment of a European Organization for Nuclear Research, member states undertake to exchange not just scientific staff but also scientific and technical information.[1]

[1] 'The history of CERN', https://home.cern/about/who-we-are/our-history.

Fast forward four decades, another world-famous example of collaboration and sharing emerges: the Human Genome Project. In little more than a decade, the United States, the UK, German, French, Japanese and Chinese researchers worked to create a DNA sequence of the entire human genome and of five other species. When it was decided in 1996 that all DNA-sequence data that the Human Genome Project generated would be released into the public domain ('Bermuda Principles'), this was seen as a radical but necessary step.

This data policy was considered essential to achieve the desired speed and quality of work, to address concerns about gene patents impeding scientific advancement and to ensure there were minimal restrictions on data access.[2] The Human Genome Project and the Large Hadron Collider research are two examples of initiatives that were large-scale, international, collaborative, and generative of large amounts of data. From that perspective it is not surprising that institutional arrangements for data sharing were made. In fields such as molecular biology and astronomy, data sharing has a long history.[3]

Today, the classic model for sharing research through journals, papers and other publications is increasingly complemented by making raw or derived data when results are published. This serves to verify outcomes but also enables re-use of data for other projects. It is not yet common for data to be released while the research is ongoing because it does not sit well with the competitive nature of scientific research and the premium put on original publications. Another factor slowing early data release is that the same data set may be used for multiple publications. Nevertheless, ex-post publication of research data under open access conditions is set to become the norm.

Such data can be held in commercial or non-profit electronic repositories, covering single or multiple institutions. Some are generalist like Figshare (commercial, Springer Nature) and Zenodo (non-profit, operated by CERN); others are domain-specific like the Endangered Languages Archive or the Polar Data Catalogue. There are thousands of repositories, set up by many different communities and companies. DataCite, the non-profit organization that provides persistent identifiers (Digital Object Identifiers or DOI) for unique data sets, already caters to some 2,300 repositories worldwide. Repositories, although key, are only one element of research infrastructures needed to build

---

[2] Kathryn Maxson Jones, Rachel A. Ankeny and Robert Cook-Deegan, 'The Bermuda Triangle: The Pragmatics, Policies, and Principles for Data Sharing in the History of the Human Genome Project' (2018) 51 *J Hist Biol* 693.

[3] See Jones and others (n 2) for a historical overview in the field of genetics; and for astronomy: Alberto Pepe and others, 'How Do Astronomers Share Data? Reliability and Persistence of Datasets Linked in AAS Publications and a Qualitative Study of Data Practices among US Astronomers' (2014) 9(8) *PLoS ONE* e104798.

the so-called science cloud.[4] This partially explains the movement towards a top-down approach.

3. THE TOP-DOWN TURN TO OPEN ACCESS

Since the launch of the Human Genome Project, its principal funders – the US National Institutes of Health (NIHS) and the UK's non-profit Wellcome Trust – have adopted the policy that research data generated in all projects they fund be shared as openly as possible. The EU has followed such a path as well. The latest multi-billion framework programme Horizon Europe (2021–2027) requires research data of funded projects to be FAIR (Findable, Accessible, Interoperable, and Reusable) (defined in more detail below) and open by default. Earlier programmes focused primarily on open access to publications.

It was during 2007 that the European Commission drafted its first policy, ultimately resulting in the 2012 Recommendation on access to and preservation of scientific information. This non-binding instrument was updated in 2018, with stricter guidelines for research data. Open science policies of governments around Europe include a similar approach to research data, but they can no longer choose their own path. By mid-2021, all EU member states must have policies in place that mandate that research data made available in repositories is also available for re-use by others, including commercial re-use. This follows from the 2019 Open Data Directive, discussed below. In the past years, the European Commission has also steadily developed a science data space, the European Open Science Cloud. This is now part of the larger EU data strategy to create 'common European data spaces' in areas such as health, agriculture, and mobility.

This shows a growing institutionalization of research data management and sharing, driven to a substantial degree by the power of external science funders. The latter have become so important that they are able to dictate when and how researchers must share. This leverage exists even though, in many countries, direct government funding of research through universities accounts for a much larger share of research funding.

4. FROM PUBLICATIONS TO DATA

At first glance, there is a clear logic to mandating open access to research data. After opening up access to publications comes opening up data and other

[4] European Commission, 'Prompting an EOSC in practice: final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC)' 2018 https://data.europa.eu/doi/10.2777/112658.

products of research. Indeed, the narrative of sharing research data mimics that of open access publishing.

The so-called Plan-S (2018) is an excellent example of that narrative, and of the outsized influence of funders on academic publishing. Funders and EC policy makers had become impatient with the snail's pace at which academic publications became open access. Universities were struggling to reach agreements with large publishers, which control back catalogues and license e-journals as part of package deals, also known as Big Deals. Pressed by the European Commission and supported by the European Research Council (ERC), a number of national research funding organizations from 12 European countries launched an ambitious plan to accelerate the transition. The stated ambition of this Coalition-S is that from 2021 'all scholarly publications… must be published in Open Access Journals, on Open Access Platforms, or made immediately available through Open Access Repositories without embargo'.[5]

The funders undertook to impose these obligations on researchers and their institutions through grant agreements. According to the plan, authors or their institutions should 'retain copyright' to their publications. Yet, at the same time, all publications are expected to be published under a Creative Commons Attribution license ('CC BY') or similar open license.

Plan-S follows the so-called Berlin Declaration that called for the rapid transition to open access publishing of academic publications and the use of Creative Commons attribution licenses. Initially conceived at an international gathering of a dozen universities, libraries, research organizations and public funders, the Berlin Declaration now has over 700 signatories.[6]

The language of both the Berlin Declaration and Plan-S clearly illustrates that a major driver behind open access movements is the wish to roll back the control that commercial publishers have gained over science publication systems. In the words of Plan-S, it is necessary to 'to regain ownership of the rules governing the dissemination of scientific information'. Publication paywalls limit access, which 'constitutes an absolute anomaly' because 'no science should be locked behind paywalls!'[7] It is also evident that copyright plays a major role.

---

[5] See https://www.coalition-s.org/about/. Research funders from the UK, Quebec (Canada) and Jordan also formally endorsed Plan-S.

[6] The Berlin declaration (https://openaccess.mpg.de/Berlin-Declaration) took inspiration from the Budapest Declaration (https://www.budapestopenaccessinitiative.org/), spearheaded by the Open Society Institute (2001), and the Bethesda Declaration (http://legacy.earlham.edu/~peters/fos/bethesda.htm), which promoted open access specifically for the biomedical community (2003).

[7] ibid.

The framing of open access as authors taking back control is however double-faced. The insistence on the use of open licenses like the Creative Commons Attribution 4.0 license[8] indeed prevents academic publishers from controlling copyright. What it does not achieve is to return any meaningful control to authors. Using CC BY means anyone in the world can use the work, for free, for any purpose, as long as the author is credited. CC Zero, also called a public domain dedication, goes further by relinquishing copyright and similar rights to the fullest extent permitted by law.

It is worth flagging that publishers continue to play a crucial role in steering author behaviour. In large part, this is due to the publish or perish nature of academia, where the individual author has no bargaining power and journals run a take it or leave it system. In the past, the publisher was able to acquire copyright from the author in lieu of publication. Now, the publisher prescribes that the author must not transfer her copyright to the publisher, but instead pay to be published and license the use of her work freely to the entire world. The one great advantage for the author is that access to her work is wider. The downside is that the author has no more control over her work than if a publisher owned the copyright. Increasingly, journals demand that the research data on which articles are based is made available in a repository, also under an open license.

Funders and libraries recognize that it is more efficient and fair to have a system where individual authors do not have to pay article processing or book production fees themselves. But buying fees off collectively is no easy task, as science publishers seek to maximize returns.[9] All the while, the publishing industry has been steadily expanding its reach to new business models around data analytics, metrics, research management and performance tools.[10]

Meanwhile, academic communities are concerned that increased dependence on commercial data analytics providers (often academic publishing conglomerates), data processing and cloud service providers (often big technology firms) threatens their autonomy and ultimately their capacity to produce

---

[8] CC-BY and CC-0 are the most commonly imposed by journals and funders, but less permissive Creative Commons Licenses also exist, see https://creativecommons.org.

[9] Transformative agreements ('publish and read' agreements) combine access to journal portfolios with immediate open access without additional payment of fees. Negotiations are tough, see e.g., in Germany https://www.projekt-deal.de/about-deal/ and in The Netherlands https://www.universiteitenvannederland.nl/en_GB/open-science-agreement-with-elsevier.html.

[10] Elsevier's parent RELX e.g., says: 'We leverage deep customer understanding, combining leading content and data sets with powerful global technology platforms, to build sophisticated analytics and decision tools that deliver enhanced value to our customers': RELX Annual Report 2021.

quality research.[11] Who controls research data, and who is able to create value from it is a matter of public interest. Surveys among researchers show that key concerns around data sharing are about the risk of misuse of openly shared data, a lack of recognition/attribution and legal uncertainty around intellectual property implications. So how do EU data policies effect research?

5. RESEARCH DATA IN THE EU DATA STRATEGY

It is important to understand the open access movement for publications because its narrative plays a major role in the drive towards open access to research data. Until recently, open access at EU level was unregulated.[12] EU copyright law itself, although largely harmonized, says little about initial ownership, transfers and copyright contract law. Some member states have provisions in their copyright contract law that allow researchers to make their works available in repositories even when the copyright is owned by publishers or others, but EU copyright law is silent on this.[13] No significant changes to the Database Directive[14] are foreseen. All that the European Commission has proposed is that Internet of Things data should somehow be excluded from attracting right under the Database Directive.[15] As it stands then, collections of research data (data sets even) might attract intellectual property.

All the same, research data governance sits in an increasingly complex web of legislation. The 2019 Open Data Directive[16] brought research data into the ambit of the general framework for re-use of public sector information, although not all its rules apply. Member states' implementation of this law is slow, but once complete, researchers can benefit from wider availability of public sector data. However, as holders of data, research performing organ-

---

[11] See eg Leiden Principles on the role of research-intensive universities as the world emerges from the COVID-19 pandemic, May 2022; 'LERU Data Statement' December 2021, https://www.leru.org/publications/is-university-autonomy-threatened-by-eu-data-policy-and-law.

[12] The text- and data-mining exception of Article 3 'DSM Directive' (Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC [2019] OJ L130/92) has no direct bearing on open access.

[13] E.g., the Dutch Copyright Act has had such a clause since 2015.

[14] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L77/20.

[15] See European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)' COM (2022) 68 final, art 7.

[16] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast) [2019] OJ L 172/56.

izations including universities and their staff also come under obligations to allow others re-use. Since it is a most important piece of legislation, it pays to have a closer look.

5.1 The Open Data Directive

The current Open Data Directive replaced the 2013 Directive on the re-use of public sector information, which in turn contained a tightening of the rules of the original 2003 Directive.[17] Its roots can be traced back to the European Commission's Guidelines for improving the synergy between the public and private sectors in the information market of 1989. These aimed to make more data produced in the public sector available for commercial exploitation by the nascent database industry. Often mentioned were legal information (legislation, court decisions), meteorological data, statistics and mapping data. It was actually the private sector who lobbied for a commercial right of access to public sector databases. The idea was there should be an EU instrument that would oblige public sector bodies to allow commercial exploitation of their information resources, and that would also ban them from providing information services that the private sector was able to offer. Publicly funded academic research and publications did not appear to be on the radar of policy makers back then and did not come fully into view for another 30 years.

The key impact of the Open Data Directive is to oblige member states to ensure that public sector bodies make the information they hold ('any content whatever its medium') available for commercial and non-commercial re-use at non-discriminatory terms. This obligation only applies to information collected or produced in the course of public tasks and to which public access is allowed, e.g., on the basis of (national) freedom of information acts. The idea that there is additional, external value to be derived from information resources created in the course of public tasks has always been central to the project. The initial focus was on the development of new commercial pan-European services it would facilitate, and the instrument was based on the EU's legislative competence to further the common market. Allowing re-use was said to serve a host of other goals as well, such as improving public service delivery and public accountability.

That the Directive now has Open Data prominently in the title shows its aspirations. In a broad sense, open data means data that 'anyone can freely access, use, modify, and share for any purpose (subject, at most, to require-

[17] Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information [2003] OJ L 345/90, as revised by Directive 2013/37/EC [2013] OJ L175/1.

ments that preserve provenance and openness)'.[18] To enable this, data must be unencumbered by intellectual property restrictions (that is, be legally open) and be technically open. The latter means it must at least be machine-readable and available in an open format. Public sector bodies must do this wherever possible.[19]

The current Directive stipulates that re-use must be allowed with as few restrictions as possible. If terms and conditions are imposed, these must be non-discriminatory (treat similar users equally), transparent and wherever possible standardized and electronically available. Recital 44 emphasizes that open licences 'should play an important role' in this respect. Also, public sector bodies are banned from exercising any *sui generis* database rights they might have to restrict re-use beyond the limits set by the Open Data Directive (art 1(6)). Of note, the Directive does not apply to documents held by public sector bodies, if third parties hold intellectual property rights in them. Another key principle is that re-use should be free; no royalties or other use charges should be levied.

The Open Data Directive only explicitly regulated academic research data in 2019. Earlier versions of the Directive contained a messy basket of exceptions while applying its terms to a broad range of public sector bodies. It is still not crystal clear which academic organizations are covered, but research branches of public universities, publicly funded research performing organizations and university libraries are included. There is now one provision specifically targeting research data. Under Article 10, publicly funded and publicly deposited research data 'shall be re-usable' under the same general provisions as other public sector data. In addition, member states must develop 'open access' policies for research data that is produced with (partial) public funding. Such policies must be in line with the Commission's revised Recommendation on access to and preservation of scientific information (2018) and be directed at domestic research organizations and funders. In a roundabout way, what started out as urgent advice from the European Commission has now become law. Like the Recommendation, Article 10 allows legitimate commercial interests, knowledge transfer activities and pre-existing intellectual property rights to be considered when crafting open access policies.

Article 10 also asks that research data be made FAIR. That acronym represents the four key characteristics of well-managed research data, which enable discovery, evaluation and downstream reuse: Data must be Findable,

[18] See for a definition of open: https://opendefinition.org/.
[19] Open Data Directive, art 5(1).

Accessible, Interoperable and Reusable.[20] These four aspects are elaborated in 11 guiding principles, drafted by a diverse group of academics, institutions and some private sector bodies (such as publishers).[21] The principles focus on technical qualities of datasets, metadata, repositories and on communication protocols and tools. In several ways, these guidelines are similar to the technical principles in the Open Data Directive aimed at easy identification and exchange of data. The Open Data Directive also sets out legal conditions, which is something FAIR does not directly address, just as it does not address ethical or moral aspects of openness.[22]

The Directive thus recognizes that it is difficult for universities to maximize the release of research data while also fulfilling the 'valorization' tasks many have, including the commercialization of research outputs through knowledge transfer to the commercial sector or state companies. There is a growing body of research on the effects of open versus proprietary dissemination of data on follow on research and innovation. In areas where data play an important role in the ability to successfully patent a process or product, data release policies could impact research institute's ability to commercialize research. On the other hand, intellectual property (and restrictive licensing terms) can also hinder follow-on research and innovation.[23]

5.2 Data Governance Act and Data Act

The Data Governance Act[24] requires public sector bodies to make available, with appropriate clearances, data that is not subject to the Open Data Directive due to data protection, commercial secrecy and third-party intellectual property constraints. Research communities are meant to be a major beneficiary.

---

[20] These elements have been identified by authors and institutions in various disciplines, but FAIR has gotten particular traction, see Mark D. Wilkinson and others, 'The FAIR Guiding Principles for Scientific Data Management and Stewardship' (2016) 3 *Scientific Data* 160018.

[21] ibid.

[22] Barend Mons and others, 'Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud' (2017) 37(1) *Information Services & Use* 54.

[23] The Human Genome Project and its private sector competitor Celera Genomics provide an interesting comparison. Celera managed to commercialize unique sequencing data (without intellectual property) because medical companies and other clients would not wait for the HGP data to become freely available. See Heidi L. Williams, 'Intellectual Property Rights and Innovation: Evidence from the Human Genome' (2013) 121(1) *Journal of Political Economy* 1–27.

[24] Regulation (EU) 2022/868 of the European Parliament and of the Council on European data governance (Data Governance Act), OJ 2022, L 152 (effective from 24 September 2023).

Further, by introducing a light regulatory regime for data altruism organizations, the EU also hopes to increase data pools for research and other public interest purposes.

The third important instrument in the EU data strategy is the (proposed) Data Act. It sets out various rules geared towards developing data spaces, such as the European Open Science Cloud. But it also creates obligations for data holders with respect to licensing terms, regardless of whether their obligation to provide access to data derives from the Data Act or any other Union legislation, including implementing laws. The concept the Data Act uses for licensing terms is FRAND: Fair, Reasonable and Non-Discriminatory terms. This is borrowed from the licensing of standard essential patents,[25] and less far-reaching than what the Open Data Directive mandates for public sector data. Under the Data Act, the data holder is free to set the terms of data access – as long as they are FRAND – but cannot contract around the provisions on data access to the detriment of the data recipient. The obligation to provide access does not force data holders to divulge trade secrets but would task the European Commission to develop non-binding model contractual terms on data access and use.

6. CONCLUSION

Access and re-use of research data is fast becoming institutionalized at the EU level. This development is driven by a combination of research funders exercising their power to impose rules, legislation implementing the EU data strategy, and the creation of a European Science cloud. Questions can be raised about whether applying generic instruments like the Open Data Directive and the proposed Data Act to (publicly) funded research is actually the right approach. The Open Data Directive is a generic instrument designed for government information, not for academic research data created in highly competitive environments that put a premium on private-public cooperation and commercialization of research outputs (technology transfers, etc.). More importantly, both the Directive and research funders push open content licenses, particularly Creative Commons Attribution licenses, as the gold standard for licensing data sets and other research products, not just academic publications. But these licenses are not well-suited to data. Making one private ordering licensing scheme a key building block of government open science policy in Europe is a curious move. Finally, the upcoming Data Act seems to

[25] On the benefits of FRAND (and other private ordering mechanisms) to counter problems of the current IP system, see Rochelle Dreyfuss, 'The Challenges Facing IP Systems: Researching for the Future' in Peter Drahos, Gustavo Ghidini and Hans Ullrich (eds), *Kritika: Essays on Intellectual Property* (vol 2, Edward Elgar 2020).

create new powers and obligations that cut through the policies of the Open Data Directive and rely on member states to enact open science policies.

Strikingly, scant attention is being paid to the impact on academic freedom of individual researchers and teams. Lip service is paid to Article 13 of the Charter of Fundamental Rights of the EU, which guarantees that 'arts and scientific research shall be free of constraint' and that 'academic freedom shall be respected'. But the rhetoric around open research data seems to primarily echo frustrations about the stranglehold that large academic publishers have had over university libraries. 'No science should be behind paywalls!' as research funders exclaimed, does not begin to address the complexities of safeguarding academic freedom in a data-driven world.