

Open Science and Data Protection: Engaging Scientific and Legal Contexts

Ludovica Paseri*

**Post-Doc Research Fellow at the Law Department,
Università degli Studi di Torino*

Abstract. This paper analyses the relationship between open science policies and data protection. In order to tackle the research data paradox of the contemporary science, i.e., the tension between the pursuit of data-driven scientific research and the crisis of repeatability or reproducibility of science, a theoretical perspective suggests a potential convergence between open science and data protection. Both fields regard governance mechanisms that shall take into account the plurality of interests at stake. The aim is to shed light on the processing of personal data for scientific research purposes in the context of open science. The investigation supports a threefold need: that of broadening the legal debate; of expanding the territorial scope of the analysis, in addition to the extra-territoriality effects of the European Union's law; and an interdisciplinary discussion. Based on these needs, four perspectives are then identified, that encompass the challenges related to data processing in the context of open science: (i) the contextual and epistemological perspectives; (ii) the legal co-ordination perspectives; (iii) the governance perspectives; and (iv) the technical perspectives.

Keywords: Open Science; Data Protection; Privacy; Research Data; Data Sharing; Scientific Research; Governance Mechanisms.

1. Introduction: The Research Data Paradox of Contemporary Science

The field of scientific research is currently experiencing a paradox: on the one hand, a considerable trend towards data-driven science is emerging (Leonelli, 2018); on the other hand, a profound crisis, the so-called “reproducibility crisis”, is taking place.

The trend towards data-driven science requires an ever-increasing volume of data (Resnik, 2005), becoming essential for implementing research projects. As a consequence, there is a growing demand for computational power and methodologies that are able to take full advantage of the elaboration of such data, in a process of technological convergence (Pagallo, Durante, Monteleone, 2017, 59).

On the other hand, however, one of the major problems in contemporary science is the so-called “crisis of reproducibility” (Baker, 2016) or, adopting another categorization, “crisis of repeatability” (Nosek *et al.*, 2022). The causes of this crisis are manifold. Difficulties may arise due

to a lack of raw data or an unwillingness from researchers to share their data (Miyakawa, 2020); a lack of sharing the code of the algorithms at the basis of the research project (Hutson, 2018); or, sometimes, the deadlock is due to the inherent difficulties of the falsifiability of the scientific content that is meant to be subjected to peer review – e.g., think about the issue of falsifiability of string theory, (Ritson, Camilleri, 2015), (Greene, 1999).

Therefore, the research data paradox of the contemporary science can be defined as the tension between the pursuit of data-driven scientific research on the one hand and, the overwhelming challenges of repeatability of such data-driven research projects and their results, on the other.

In light of the current paradox, there are several issues of regulation (or lack of regulation) worth analysing. This contribution draws the attention to a set of issues which are often overlooked, namely, the interplay between open science policies and the protection of personal data processed for scientific research purposes (Pagallo and Bassi, 2013).

The fundamental relevance of the topic stems from the fact that, in the last five years, open science has shifted from being a movement supported by a part of the scientific community to being a fully-fledged policy institutionalised by the European Union, national and international actors (Paseri, 2021, 165-166). The European institutions have chosen open science as the default approach for research funded by the new programme Horizon Europe¹. In parallel, in the United States, 2023 is identified as the “Year of Open science” (The White House, 2023). In addition, in 2021 UNESCO released the first Recommendation on open science (UNESCO, 2021), aiming to monitor the progress of the openness in the scientific research in every area of the world. In this significant scientific transformation, UNESCO’s role is to promote the local peculiarities of each community, ensuring that no one is left behind.

These initiatives illustrate how much has happened in the last years: Open science is no longer only synonymous with open access to the scientific literature – it is much more than that – nor is it the bottom-up instances of the scientific community. Indeed, today, open science is an umbrella term that could best be understood as open scientific research process, in which the principles – i.e., openness, cooperation, inclusivity (Leonelli, 2023), collaboration, sharing, independence, integrity and

¹ Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013 (Text with EEA relevance), ELI: <http://data.europa.eu/eli/reg/2021/695/oj>.

transparency – encompass both the inputs of the process (data and research funding) as well as the outputs of the process (publications, educational resources, conferences and dissemination activities, etc.), also engaging the actors, instruments and methodologies of research. The open science, now, is the approach that aims to open up every phase of scientific research, involving in this collaborative process a wider range of actors at different stages (Paseri, 2022a).

Alongside such policies that aim to promote science “as open as possible, as closed as necessary”, attention must be drawn to the legal framework on the processing of personal data for scientific research purposes, which in the European Union is represented by the General Data Protection Regulation (GDPR)² and its national, adapting legislative provisions.

The investigation of the relationship between open science and personal data protection is driven by a threefold need: (i) broadening the legal debate; (ii) expanding the territorial scope of analysis beyond the European Union; and (iii) fostering an interdisciplinary discussion.

(i) Broadening the Legal Debate

The legal debate related to open science and policies for its implementation has been flourishing for years in the area of Intellectual Property (Guibault, 2013), (Peters and Margoni, 2016), (Caso, 2019), (Willinsky, 2022). Admittedly, there are major knots to be untangled in that domain and important battles are waging in the tension between open and closed science.

However, the great role played by personal data in scientific research – think about the COVID-19 pandemic (Besançon *et al.*, 2021) – calls for a broadening of the legal debate that engages the field of privacy law, data protection law and ethics.

(ii) Expansion of Territorial Scope

There is a need to broaden the debate with regard to geographical scope. First, the GDPR, under Article 3 outlining the territorial scope, lays the groundwork for what has been identified as the extraterritoriality of the GDPR (Greze, 2019), stating that the Regulation “applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of

² Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), ELI: <http://data.europa.eu/eli/reg/2016/679/oj>.

whether the processing takes place in the Union or not” (Article 3.1 GDPR).

In addition, this need to adopt a global perspective in addressing the link between open science and data protection is due to the inherent global nature of science. Although from a legal point of view, the scientific research regulatory framework is traditionally developed on national basis, science *per se* transcends national borders and, indeed, this global dimension is all the more triggered by the digital revolution and the potential it offers.

(iii) Interdisciplinarity

The complexity of the topic and the multiple, different, and sometimes conflicting interests at stake plead for an interdisciplinary approach. The legal debate on the relationship between open science and data protection requires the involvement of several fields of knowledge: legal, from different branches of law, ethical and philosophical, technical and sectoral.

In light of this threefold need, the JOAL special issue on “Open Science and Data Protection” is divided into two parts. The JOAL special issue vol. 11, no. 1, hosts contributions from authors with different backgrounds: philosophical, administrative, economic, health-related. The JOAL special issue vol. 11, no. 2 gathers legal and policy contributions, from philosophy of law, private law, public law and comparative law. The perspective of this JOAL special issue is European, looking primarily at the relationship between EU open science policies and GDPR. However, in light of this need to expand the debate from a territorial perspective, the JOAL special issue also embeds the US and Australia standpoint.

The paradox of research data and the threefold need (i.e., broadening the legal debate; expanding the territorial perspective; and adopting an interdisciplinary approach) underlie and drive the investigation of this JOAL special issue concerning the relationship between open science and data protection: is there a possible convergence? The second section explores the reasons supporting a potential convergence. The third paragraph looks at the outstanding issues, presenting the approach adopted in this JOAL special issue to frame the challenges. Finally, the fourth paragraph concludes the analysis pointing out some possible future strands of research.

2. Data Protection and Open Science: A Possible Convergence?

The relationship between the data protection regime and the open science policies is often described as a clashing tension between the *openness* promoted by the new scientific approach and the *closure* imposed by the regulatory framework. Over the years, there have been quite a few stances that have identified the GDPR as a real limitation for open science or as a barrier for global science (Phillips and Knopper, 2019), (Eiss, 2023).

While it cannot be denied that there are some knots to be untangled, nevertheless, from a theoretical perspective, the alleged disagreement is based on a twofold flawed premise (section 2.1), is not confirmed by empirical data (section 2.2) and, moreover, and is misleading (section 3.3).

2.1. A TWOFOLD FLAWED PREMISE

The clashing conflict between openness and data protection rests on two erroneous premises concerning, on the one hand, a misinterpretation of the concept of openness conveyed by the open science approach and, on the other hand, an incomplete view of the objective pursued by the GDPR.

The openness promoted by the open science policies is not indiscriminate but always the result of a balancing of the interests at stake. Moreover, the primary objective currently pursued by open science is related to the awareness of the change taking place and the consequent reshaping of processes, practices and management of the research sector. In other words, open science is the approach through which a greater transparency of the scientific research process is fostered: In the long run it may even lose the adjective ‘open’, to be “simply science” (Watson, 2015).

In relation to research data, this goal therefore results chiefly in the promotion of good management of research data. Such sound management can be achieved in many ways. Among the various approaches, the FAIR principles, which aim to harmonise data management by means of common guidelines, stand out, providing research data that are findable, accessible, interoperable and – possibly – reusable (Wilkinson *et al.*, 2016). Having accessible research data does not mean share data indiscriminately, without control or security. In a nutshell, findability has to do with the long-term preservation of data and datasets, for instance, by attributing a unique identifier to each resource. Accessibility means the possibility of – potentially – accessing research data:

this means storing them in a repository that is either institutional, i.e., of the university or research centre where the research is conducted, or relevant and *trusted* in the scientific field of reference. The accessibility of FAIR principles, in other words, intends to affirm accurate archiving practices of research data and does not indicate indiscriminate access to data – perhaps personal – by whoever. The aim, rather, is to contrast bad practices, unfortunately widespread, of research data stored in proprietary clouds or worse, on researchers’ private hardware, of which there is no way of keeping track or avoiding potential data breaches. Interoperability, then, is “the ability of data or tools from non-cooperating resources to integrate or work together with minimal effort” (Wilkinson *et al.*, 2016, 3). Finally, reusability means the set of descriptive features that each dataset must have to make those who did not participate in the collection or creation of the data understand what they can legitimately do with it. Such a description also implies the identification of the suitable licence, which brings legal certainty regarding the legitimate uses of that data, before providing the access to it.

On the other hand, the second mistaken premise is the interpretation of the GDPR and the legal framework concerning the processing of personal data as a limitation for the data sharing. The GDPR is not solely aimed at the protection of personal data. Article 1 of the GDPR, under the heading “Subject-matter and objectives” in paragraph 2 states that the Regulation “protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data”. However, the following paragraph 3 emphasises that the “free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data”. It is worth mentioning that the GDPR is a piece of legislation that the European Union has envisaged as a key component of the Digital Single Market strategy. In 2015, the European Commission defined the Digital Single Market as follows:

A Digital Single Market is one in which the free movement of goods, persons, services and capital is ensured and where individuals and businesses can seamlessly access and exercise online activities under conditions of fair competition, and a high level of consumer and personal Data protection, irrespective of their nationality or place of residence. Achieving a Digital Single Market will ensure that Europe maintains its position as a world leader in the digital economy, helping European companies to grow globally.³

³ European Commission, *Communication on A Digital Single Market Strategy for Europe*, COM/2015/0192 final, p. 3, 2015.

In other words, “the economic exploitation of data requires the creation of a digital single market that provides the best conditions for the free circulation of data, allowing their collection, dissemination, aggregation, and so on” (Durante, 2019, 130). The current EU legal framework emerges precisely from this context: “The creation of this market in turn requires establishment of a framework providing legal certainty, as a prerequisite for economic investment, innovation and development of business” (Durante, 2021, 130). This dimension should not be bypassed when investigating the relationship between open science and data protection, *a fortiori* since the processing carried out in this context pursues scientific research purposes.

2.2. LACK OF EMPIRICAL EVIDENCE FOR CONFLICT

The assumed conflict between openness and data protection is not empirically proven. Consider that in the Eurobarometer analysis on the impact of digital technology on the everyday life of individuals, released in March 2020, “the majority of respondents said they would share their data mainly to improve research and medical care” (Pagallo, 2022, 75).

A trend towards sharing data for research purposes emerges, so much so that some academics have defined the tension between data sharing and data protection as “anecdotal and empirically unjustified” (Ienca, 2023, 2). A recent study conducted in Switzerland in 2022 (Pletscher *et al.*, 2022) illustrates that “survey results show that although privacy and data protection concerns are very common among the Swiss population (74%), the large majority (71%) of respondents (with peaks of 81% among people with chronic diseases) reported that they are nevertheless willing to share their data for medical research” (Ienca, 2023, 2).

The study, therefore, clarifies that data sharing for research purposes is not hindered by data protection law: on the contrary, this is the guarantee underlying the fiduciary covenant between individuals and researchers. The legal framework on data protection is perceived as the set of provisions safeguarding individuals who choose to share their data and foster their re-use for research purposes. Rather, a clashing tension arises between the protection of personal data and a *closed* science that does not respect the principle of transparency and integrity, avoiding providing information on its research data management, hampering the repeatability of scientific experiments in disregard of the scientific method.

2.3. A MISLEADING REPRESENTATION

Describing the relationship between open science and data protection in terms of a barrier or limit does not clearly frame the issues at stake and is therefore not helpful in finding adequate solutions to tackle the problems. Some of the criticisms of those who argue for this irreconcilable tension are well-founded. For instance, consider the risks of re-identification of anonymised personal data processed for research purposes (Erb *et al.*, 2021, 3). However, a checks and balances approach is needed to meet these challenges. On the one hand we have the right to the protection of personal data, as enshrined in the Article 8 of the Charter of Fundamental Rights of the European Union. On the other hand, we have the right to science, with all its levels of protection (Paseri, 2022b, 518); (Porsdam and Porsdam Mann, 2020), (Perrone, 2020). Admittedly, in balancing these two rights, several complex issues arise: They need to be addressed resulting in “the continual management of boundaries between different spheres of action and degrees of disclosure within those spheres. Boundaries move dynamically as the context changes. These boundaries reflect tensions between conflicting goals; boundaries occur at points of balance and resolution” (Palen, *et al.*, 2003, 131). In other words, the interplay between open science and data protection becomes a matter of design: In order to tackle the current challenges, it is crucial to design mechanisms (Dennis *et al.*, 2019, 1843) capable of taking all the interests into consideration at the governance level.

In addition, this balancing of interests is realised in a complex and fragmented legal framework represented by the provisions provided by the GDPR and those established by the various national regulations on scientific research. In this regard, it is essential to refer to the Article 89 of the GDPR, titled “Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes”. Paragraph 1 states that the processing of personal data for scientific research purposes, as well as for archiving, historical research and statistical purposes must provide adequate guarantees for the rights and freedoms of the data subject, i.e., the identified or identifiable person to whom the personal data processed pertain, by providing technical and organisational measures. In identifying such measures to ensure the rights and freedoms of the individual, central importance is given to the principle of minimisation of processing, according to which the personal data processed must be “adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed”, as enshrined in Article 5(c) of the GDPR.

Starting from the guarantees required by the Article 89 of the GDPR, however, the EU law on data protection provides for a derogatory discipline for the processing of personal data for scientific research purposes⁴: On the one hand, the GDPR itself provides for a number of specific exceptions, in various sections of the normative text; on the other hand, broad leeway is left to national legislators in this specific area. In fact, in paragraphs 2 and 3 of the Article 89 of the GDPR, the European lawmaker establishes that for the processing of personal data for scientific research purposes, both the EU law and the national law of the Member States may provide for a set of exceptions to a number of requirements set out in the GDPR. The European lawmaker, therefore, specifically allocates a certain scope of national autonomy, regarding the sector of scientific research (Ducato, 2020). This fragmented European framework in the field of scientific research increases the legal uncertainty.

It seems fair to admit that the traditional perspective of the relationship between open science and data protection in terms of an unbridgeable conflict leaves no room for any potential convergence. By contrast, the JOAL special issue on “Open Science and Data Protection” investigates the conditions that make such convergence possible. The following section describes the approach adopted in this JOAL special issue to address the challenges of data protection in the context of open science and introduces the contributions of the authors.

3. A Multidimensional Analysis between Openness and Protection

The JOAL special issue on open science and data protection is structured in two parts. The first (vol. 11 no. 1) collects contributions that provide the necessary background to grasp the main aspects of the issues at stake. This first part is concluded by Prof. John Willinsky’s comments. Prof. Willinsky stresses the need to avoid the rhetoric that often accompanies the analysis on open science policy for “effectively competing for attention in making the case for research’s priorities” (Willinsky, 2023, 12).

The second part (vol. 11 no. 2) includes legal contributions that are closely related to the legal challenges of implementing open science policies. This second part ends with final remarks by Veronique Ciminà, who offer her insights and expertise after years of work in the European

⁴ On what is meant by “scientific research” under EU data protection law, see: (Paseri, Varrette and Bouvry, 2021, 129-130).

institutions, specifically involved in the protection of personal data. The analysis she proposes on processing for scientific research purposes in the framework of open science emphasises that “it would be inaccurate and simplistic to conclude that the data protection legal framework and open science are in contrast with one another” (Ciminà, 2023, 11).

From the overall picture of the contributions to this special issue, I identify four major challenges, adopting an interdisciplinary approach: contextual and epistemological (section 3.1); on legal co-ordination (section 3.2); on governance (section 3.3); and technical (section 3.4).

3.1. THE CONTEXTUAL AND EPISTEMOLOGICAL CHALLENGES

Exploring contextual and epistemological perspectives is crucial to be able to assess the current state of implementation of open science policies, as well as to identify frictions with the legal framework on data protection. The contribution written by Elena Giglia, under the title “Open? The Only Way Forward for Science” aims to clarify “the reasons underlying the need to foster as much as possible the sharing and re-use of research data as well as their FAIRness” (Giglia, 2023, 1). Almost as a manifesto, representing the initial bottom-up drive of open science, the author insists on the value of the FAIRness of data to generate accurate scientific research of high value that respects the principle of integrity of science.

Alongside this position paper, David Resnik offers an analysis of the concept of openness. In his contribution “Openness in Scientific Research: A Historical and Philosophical Perspective”, the author proposes an *excursus* on the notion of openness in scientific research, adopting a historical perspective, which lays the foundation for his argument in support of openness in research as the “keystone of scientific ethics and policy” (Resnik, 2023, 6).

In addition, in looking at the context, a key factor of the open science should not be forgotten: For the EU institutions, policies in support of openness in scientific research find their drive in the digital revolution. In recent years, European open science policies have been developed in close connection with policies supporting innovation and technology. This includes a strand dealing with digital sovereignty. Luc Soete and Jean-Claude Burgelman, in their contribution titled “Reconciling Open Science with Technological Sovereignty: Can the European Union do it?” question whether and to what extent the notion of openness “can be maintained as a core characteristic of European values in a world in which the geo-political tensions [...] have taken their toll” (Soete and Burgelman, 2023, 4). The interpretation of the notion of openness – mostly when conditioned by policies striving for digital sovereignty –

has a considerable impact on the research data from which innovation is engendered.

3.2. THE LEGAL CO-ORDINATION CHALLENGES

From a legal point of view, there is considerable legal uncertainty and fragmentation concerning personal data processed for scientific research purposes⁵. This uncertainty is generated, as seen above, by the highly complex legal framework. First, at the European level the GDPR and the several national adaptation law need to be taken into account. In addition, however, it should not be forgotten that the legal framework for personal data protection was not adopted from scratch, in a regulatory vacuum. On the contrary, there are many regulatory systems with which the data protection framework has to interface. In particular, Giorgia Bincoletto proposes an analysis concerning the relationship between the data protection regime, represented by the GDPR, and the European open data regime, represented by Directive (EU) 2019/1024⁶, in which the Article 10 is specifically dedicated to research data. The author focuses on the emblematic case represented by the processing of health data, arguing how “the application of a data protection by design approach on a case-by-case basis (to be preferred to a one-size-fits-all solution) allows data management practices that open the collected personal health data for specific scientific projects” (Bincoletto, 2023, 20).

Then, Dara Hallinan, Franziska Boehm, Annika Külpmann, and Malte Elson present a contribution titled “(Un)informed Consent in Psychological Research: An Empirical Study on Consent in Psychological Research and the GDPR”, that deals with the role of informed consent as a legal basis for the processing of personal data for scientific research purposes in the field of psychology. The authors illustrate the difficulties associated with the provision of consent to the processing of data that proves not to be a suitable legal basis (Hallinan *et al.*, 2023, 21). *A fortiori*, the provision of consent represents a problematic aspect in an open science context, which aims first and foremost at a good management of scientific research data. As a consequence, safeguarding data subjects and providing them with adequate information about processing activities becomes a cornerstone of data stewardship and FAIRness of research data, key factors of the open science approach.

⁵ EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space, 2022, p. 7.

⁶ 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), ELI: <http://data.europa.eu/eli/dir/2019/1024/oj>.

Furthermore, concerning the issue of legal co-ordination, Valentina Colcelli and Roberto Cippitani, with their analysis on “Circulation of personal data and non-personal data within the European Research Area for research and health purposes”, take into account the soft law level and the recent European policy developments presented by the European Commission. Specifically, on the one hand they investigate the relationship between the European data strategy and the European Research Area (ERA), on the other hand, look at recent proposal of the European Health Data Space (EHDS) (Colcelli and Cippitani, 2023).

3.3. THE GOVERNANCE CHALLENGES

Since many levels, actors, and systems participate in the open science scenario, several stakeholders are involved in the processing of personal data, pursuant to the broad definition of processing activities provided by the Article 4 of the GDPR. For this reason, it is crucial to wonder how such different levels, actors, and systems do interact and what are the models of governance currently adopted. In particular, Anna Berti Suman addresses the management and governance of research data in a specific situation: the crisis scenario. In her contribution “Citizen-Gathered Data to Support Public Services Under Emergencies: Promises and Perils of Openness” the author examines the role of open access to research data that fall into the category of “citizen-generated data”, questioning the notion of “technology appropriation” and “collective intelligence” (Berti Suman, 2023, 6) in the context of citizen science initiatives.

Moreover, the scientific research projects can frequently be conducted between research entities even far apart from each other and this thanks to the facilities and opportunities that digital technologies offer. This required extending the scope of the study. In particular, Roxanne Missingham offers an analysis of Australian open science and data protection policies. In her contribution under the title “Policy and Legislation Challenges for Open Science: Developments in Australia”, the author describes recent developments in privacy and research data management (Missingham, 2023).

Anat Lior, adopting the US perspective, in her paper “Private and Academic AI Collaboration: Opportunities and Challenges to Open Science in the US” proposes an overview of the benefits and risks of collaboration between public and private actors in the field scientific research on the Artificial Intelligence (AI) (Lior, 2023). As a result of the profound public-private interplay in an open science context, the management of personal data acquires considerable relevance, both from an ethical-legal and economic point of view.

3.4. THE TECHNICAL CHALLENGES

When investigating the challenges of processing data for scientific research purposes, the technical dimension cannot be overlooked.

In this regard, Shalini Kurapati and Luca Gilli, in their contribution titled “Synthetic Data: A Convergence Between Innovation and GDPR” explore “the role that synthetic data could potentially play in generating a convergence between the protection of the fundamental right to personal data protection on the one hand and innovation and data sharing on the other” (Kurapati and Gilli, 2023, 1). Synthetic data are included in the so-called “Privacy-enhancing technologies” (PETs): the advantages for data sharing, privacy and data protection are potentially very high. However, as pointed out by the authors, an underlying lack of communication between the legal and technical domains, combined with the legal uncertainty that *de facto* characterises the use of synthetic data, still stands as a limitation to a maximum exploitation of the potential advantages.

Then, Hammam Abu Attieh, Anna Haber, Felix Nikolaus Wirth, Benedikt Buchner and Fabian Prasser present an analysis titled “Enabling Open Science in Medicine Through Data Sharing: An Overview and Assessment of Common Approaches from the European Perspective” which focuses on the approach they have adopted in carrying out research projects involving the processing of biomedical and health data. The authors offer an overview of the different methods used for sharing biomedical data, in order to discuss their technical properties and the related legal challenges, developing their assessment on the “Five Safes Framework” (Attieh *et al.*, 2023, 3).

4. Conclusions

The JOAL special issue on “Open Science and Data Protection” that I edited intends to shed light on the processing of personal data for scientific research proposes in the context of open science. The investigation is driven by a threefold need: (i) broadening the legal debate; (ii) expanding the territorial scope of analysis beyond the European Union; and (iii) fostering an interdisciplinary discussion.

This paper claimed that, in order to tackle the research data paradox of the contemporary science, i.e., the tension between the pursuit of data-driven scientific research and the crisis of repeatability or reproducibility of science, is necessary to adopt a theoretical perspective that envisages a potential convergence between open science and data protection. This convergence is based on the identification of governance mechanisms that take into account the plurality of interests

at stake. The paper argued how believing in an unbridgeable conflict between open science and data protection is (i) based on a twofold flawed premise, (ii) is not confirmed by empirical data and, moreover, and (iii) is misleading, neither helping to tackle problems *de iure condito* nor promoting solutions *de iure condendo*.

From the renewed theoretical framework presented in section 3, four perspectives are then identified to which the challenges related to data processing in the context of open science can be drawn: (i) contextual and epistemological perspectives; (ii) legal co-ordination perspectives; (iii) governance perspectives; and (iv) technical perspectives. For each strand of investigation, the corresponding contributions of the special issue have been introduced.

The issue of open science policy implementation and the related challenges of sharing research data are very complex, involving many fields of knowledge and society at large. In addition, starting with the defined European open science framework, there is currently a great deal of turmoil at national level and many Member States are developing their own plans and strategies.

Furthermore, in the domain of personal data protection, the new adequacy decision of the European Commission “implementing decision pursuant to Regulation (EU) 2016/679 of the European Parliament and of the Council on the adequate level of protection of personal data under the EU-US Data Privacy Framework”⁷ was released on 10 July 2023. The highly anticipated decision will necessarily have a considerable impact on science, in the sharing of research data. Consider that point 11 explicitly refers to research data and the Annex I, Section II.1.b. Supplemental Principle 14 (Annex I, Section III.14.b and c) lays down specific provisions for the processing of personal data in the context of health research and clinical trials.

Another aspect worth monitoring is the so-called data altruism mechanism as regulated by the Data Governance Act (DGA)⁸, which will enter into force on 23 September 2023. The implementation of data altruism involves the Member States and entails inevitable implications for scientific research, which is specifically mentioned as a main case in the Article 2(16) of the DGA.

⁷ Commission Implementing Decision (EU) 2022/254 of 17 December 2021 pursuant to Regulation (EU) 2016/679 of the European Parliament and of the Council on the adequate protection of personal data by the Republic of Korea under the Personal Information Protection Act (notified under document C(2021) 9316) (Text with EEA relevance).

⁸ Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA relevance), ELI: <http://data.europa.eu/eli/reg/2022/868/oj>.

In light of this complexity and the ongoing technical, governance and legal developments, the JOAL special issue on “Open Science and Data Protection” is intended to be a starting point for further investigation.

Acknowledgements

I would like to acknowledge Dr. Ginevra Peruginelli and Dr. Enrico Francesconi for supporting me during this project. I would like to extend my sincere thanks to the authors and the many reviewers involved in the JOAL special issue on “Open Science and Data Protection” for developing an insightful *open* dialogue of mutual and authentic exchange and enrichment, as open science should be. I am in particular very grateful to Prof. John Willinsky, who guided me in outlining the structure of the special issue, with whom I discussed many aspects and learned a lot.

References

- Attieh, H. A., Haber, A., Wirth, F. N., Buchner, B., Prasser, F. (2023), *Enabling Open Science in Medicine Through Data Sharing: An Overview and Assessment of Common Approaches from the European Perspective*, in Paseri, L. (ed.), “Special issue on Open Science and Data Protection”, Journal of Open Access to Law, 2023, vol. 11, no. 2, pp. 1-10.
- Baker, M. (2016), *Reproducibility crisis*, Nature, 533.26, pp. 353-366.
- Berti Suman, A. (2023), *Citizen-gathered data to support public services under emergencies: promises and perils of openness*, in Paseri, L. (ed.), “Special issue on Open Science and Data Protection”, Journal of Open Access to Law, 2023, vol. 11, no. 2, pp. 1-24.
- Besançon, L. et al. (2021), *Open science saves lives: lessons from the COVID-19 pandemic*, BMC Medical Research Methodology, No. 1, pp. 1-18.
- Bincoletto, G. (2023), *Scientific research processing health data in the European Union: data protection regime vs. open data*, in Paseri, L. (ed.), “Special issue on Open Science and Data Protection”, Journal of Open Access to Law, 2023, vol. 11, no. 2, pp. 1-24.
- Caso, R. (2019), *La Rivoluzione Incompiuta. La Scienza Aperta Tra Diritto D'autore e Proprietà Intellettuale*, Ledizioni, Milano.
- Ciminà, V. (2023), *The protection of personal data for the purpose of scientific research and the Open Science framework: Ongoing challenges and future opportunities*, in Paseri, L. (ed.), “Special issue on Open Science and Data Protection”, Journal of Open Access to Law, 2023, vol. 11, no. 2, pp. 1-15.
- Colcelli, V., Cippitani, R. (2023), *Circulation of personal data and non-personal data within the European Research Area for research and health purposes*, in Paseri, L. (ed.), “Special issue on Open Science and Data Protection”, Journal of Open Access to Law, 2023, vol. 11, no. 2, pp. 1-24.
- Dennis, S. et al. (2019), *Privacy versus open science*, Behavior research methods, 51, pp. 1839-1848.

- Ducato, R. (2020), *Data protection, scientific research, and the role of information*, Computer Law & Security Review, 37, pp. 105412.
- Durante, M. (2021), *Computational power: the impact of ICT on law, society and knowledge*, Routledge, New York-London.
- Eiss, R. (2023), *We must dismantle the barriers that GDPR creates for global science*, Financial Times, 3 January 2023. <https://www.ft.com/content/622e8097-ef6b-4ec9-939e-5d95743532c1>.
- Erb, B. *et al.* (2021), *Emerging privacy issues in times of open science*, Psyarxiv.
- Giglia, E. (2023), *Open? The Only Way Forward for Science*, in Paseri, L. (ed.), “Special issue on Open Science and Data Protection”, Journal of Open Access to Law, 2023, vol. 11, no. 1, pp. 1-12.
- Greene, B. (1999), *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*, New York, W.W. Norton.
- Greze, B. (2019), *The extra-territorial enforcement of the GDPR: a genuine issue and the quest for alternatives*, International Data Privacy Law, 9.2, pp. 109-128.
- Guibault, L. (2013), *Licensing research data under open access conditions under European law*, In: Access to Information and Knowledge: 21st Century Challenges in Intellectual Property and Knowledge Governance, Edward Elgar, Cheltenham.
- Hallinan, D., Boehm, F., Külpmann, A., Elson, M. (2023), *(Un)informed consent in Psychological Research: An empirical study on consent in psychological research and the GDPR*, in Paseri, L. (ed.), “Special issue on Open Science and Data Protection”, Journal of Open Access to Law, 2023, vol. 11, no. 2, pp. 1-28.
- Hutson, M. (2018), *Artificial intelligence faces reproducibility crisis*, American Association for the Advancement of Science, pp. 725-726.
- Ienca, M. (2023), *Medical data sharing and privacy: a false dichotomy?*, Swiss Medical Weekly, 153(1), 1-4.
- Kurapati, S., Gilli, L. (2023), *Synthetic data: A convergence between Innovation and GDPR*, in Paseri, L. (ed.), “Special issue on Open Science and Data Protection”, Journal of Open Access to Law, 2023, vol. 11, no. 2, pp. 1-12.
- Leonelli, S. (2023), *Philosophy of Open Science*, Cambridge University Press, Cambridge.
- Leonelli, S. (2018), *La Ricerca Scientifica Nell'era Dei Big Data*, Maltemi, Milano.
- Lior, A. (2023), *Private and Academic AI Collaboration: Opportunities and Challenges to Open Science in the US*, in Paseri, L. (ed.), “Special issue on Open Science and Data Protection”, Journal of Open Access to Law, 2023, vol. 11, no. 2, pp. 1-14.
- Missingham, R. (2023), *Policy and legislation challenges for Open Science: Developments in Australia*, in Paseri, L. (ed.), “Special issue on Open Science and Data Protection”, Journal of Open Access to Law, 2023, vol. 11, no. 1, pp. 1-22.
- Miyakawa, T. (2020), *No raw data, no science: another possible source of the reproducibility crisis*, Molecular brain, 13.1, pp. 1-6.
- Nosek, B. *et al.* (2022), *Replicability, robustness, and reproducibility in psychological science*, Annual review of psychology, 74, pp. 719-748.
- Pagallo, U., Bassi, E. (2013), *Open data protection: Challenges, perspectives, and tools for the reuse of PSI*, in Hildebrandt, M., *et al.* (eds.), “Digital enlightenment yearbook 2013”, IOS Press, Amsterdam, pp. 179-189.
- Pagallo, U., Durante, M., Monteleone, S. (2017), *What is new with the internet of things in privacy and data protection? Four legal challenges on sharing and control in IoT*, In Leenes, R., *et al.* (eds.), “Data protection and privacy:(In) visibilities and infrastructures”, Springer, Cham, pp. 33-73.

- Pagallo, U. (2022), *Il dovere alla salute: Sul rischio di sottoutilizzo dell'intelligenza artificiale in ambito sanitario*, Mimesis, Milano-Udine.
- Palen, L., Dourish, P. (2003), *Unpacking 'Privacy' for a Networked World*, In: Cockton, G., Korhonen, P. (eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, ACM Press, pp. 129-136.
- Pasero, L. (2022a), *Accessibilità al sapere*, in Durante, M., Pagallo, U. (eds.), "La politica dei dati, Il governo delle nuove tecnologie tra diritto, economia e società", Mimesis, Milano-Udine, pp. 141-162.
- Pasero, L. (2022b), *From the Right to Science to the Right to Open Science. The European Approach to Scientific Research*, *European Yearbook on Human Rights*, Intersentia, 2022, pp. 515-541.
- Pasero, L. (2021), *COVID-19 Pandemic and GDPR: When Scientific Research becomes a Component of Public Deliberation*, in Hallinan, D., Leenes, R., De Hert, P. (eds.), "Data Protection and Privacy, Volume 14: Enforcing Rights in a Changing World", London, Bloomsbury Publishing, pp. 157-185.
- Pasero, L., Varrette, S., Bouvry, P. (2021), *Protection of Personal Data in High Performance Computing Platform for Scientific Research Purposes*, In N. Gruschka, L.F.C. Antunes, K. Rannenberg, P. Drogkaris (eds.), "Privacy Technologies and Policy. APF 2021", Springer, Cham, pp. 123-142.
- Perrone, G. (2020), *Scienza e diritti economici, sociali e culturali: Il Commento generale n. 25 del Comitato dei diritti economici, sociali e culturali*, *Diritti umani e diritto internazionale*, 3, pp. 786-795.
- Peters, D., Margoni, T. (2016), *Creative Commons Licenses: Empowering Open Access*, SSRN.
- Phillips, M., Knoppers, B. M. (2019), *Whose Commons? Data Protection as a Legal Limit of Open Science*, *The Journal of Law, Medicine & Ethics*, 47(1), 106-111.
- Pletscher, F., et al. (2022), *Willingness to share anonymised routinely collected clinical health data in Switzerland: a cross-sectional survey*, *Swiss Medical Weekly*, 152, 2324.
- Porsdam, H., Porsdam Mann, S. (eds.) (2021), *The Right to Science: Then and Now*, Cambridge, CUP.
- Resnik, D. (2023), *Openness in Scientific Research: A Historical and Philosophical Perspective*, in Pasero, L. (ed.), "Special issue on Open Science and Data Protection", *Journal of Open Access to Law*, 2023, vol. 11, no. 1, pp. 1-16.
- Resnik, D. (2005), *The ethics of science: An introduction*, Routledge, New York.
- Ritson, S., Camilleri, K. (2015), *Contested boundaries: The string theory debates and ideologies of science*, *Perspectives on Science*, 23.2, pp. 192-227.
- Soete, L., Burgelman, J.-C. (2023), *Reconciling Open Science with Technological Sovereignty: Can the European Union do it?*, in Pasero, L. (ed.), "Special issue on Open Science and Data Protection", *Journal of Open Access to Law*, 2023, vol. 11, no. 1, pp. 1-16.
- The White House (2023), *Fact Sheet: Biden-Harris Administration Announces New Actions to Advance Open and Equitable Research*, 11 January 2023, <https://www.whitehouse.gov/ostp/news-updates/2023/01/11/fact-sheet-biden-harris-administration-announces-new-actions-to-advance-open-and-equitable-research>.
- UNESCO (2021), *UNESCO Recommendation on Open Science*, pp. 1-34, <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>.
- Watson, M. (2015), *When will 'open science' become simply 'science'?* *Genome biology*, 16(1), pp. 1-3.

- Wilkinson, M. D. *et al.* (2016), *The FAIR Guiding Principles for scientific data management and stewardship*, *Scientific data*, 3(1), pp. 1-9.
- Willinsky, J. (2023), *Part I Commentary: On Not Taking Open for Granted*, in Paseri, L. (ed.), “Special issue on Open Science and Data Protection”, *Journal of Open Access to Law*, 2023, vol. 11, no. 1, pp. 1-20.
- Willinsky, J. (2022), *Copyright’s broken promise: how to restore the law’s ability to promote the progress of science*, Cambridge, Massachusetts, MIT Press.