



**HAL**  
open science

# Open science, FAIR data: Challenges and principles of opening research data in social sciences

Héloïse Berkowitz, Hélène Delacour

## ► To cite this version:

Héloïse Berkowitz, Hélène Delacour. Open science, FAIR data: Challenges and principles of opening research data in social sciences. *L'action publique des données & Les données de l'action publique - ESR*, 2022, Strasbourg, France. hal-03819303

**HAL Id: hal-03819303**

**<https://hal.science/hal-03819303v1>**

Submitted on 25 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

## EDITORIAL

# Opening Research Data: What Does It Mean for Social Sciences?

Héloïse Berkowitz<sup>1\*</sup> and Hélène Delacour<sup>2</sup>

<sup>1</sup>CNRS, LEST, Aix-Marseille Université, Aix-en-Provence, France

<sup>2</sup>Université de Lorraine, CEREFIGE, F-54000 Nancy, France

## Abstract

Recent international trends demonstrate multilevel efforts to 'open' science across its whole ecosystem and lifecycle – from capturing research data through to publishing results. In social sciences, the publication process is already largely 'open access' or transitioning toward it. However, opening research data raises specific issues and concerns for the field. Here, we set out to understand what open research data mean for social sciences, and if, why, and how data should be made open. We argue that while the ecosystem of actors, infrastructures, standards, and principles is starting to take structure in France and abroad, there are several barriers to the process of opening data in social sciences: (1) a misperception of the motivations for opening data (i.e., focusing on risks of exercising control over researchers and their academic freedom and overlooking motivations like data patrimonialization, pooling and potential synergies, trust-building, and broader engagement), (2) a system based on competition and the dominant process of 'starification' in research, (3) a lack of resources and capabilities that might further exacerbate inequalities among genders, communities, institutions, and countries, and (4) the potential risks inherent to opening data and the specific constraints posed by social science data. Against this backdrop, we investigate several ways forward to operationalize not only FAIR (Findable, Accessible, Interoperable and Reusable) but also CARE (Collective benefit, Authority to control, Responsibility, Ethics) principles for open data in social sciences, before going on to present M@n@gement's new open data policy.

**Keywords:** *Open research data; Open science; FAIR principles; CARE principles; Social sciences*

In the space of less than a decade, science has seen several deep transformations in its institutional and regulatory landscape. The European General Data Protection Regulation (GDPR) adopted in 2016 and implemented in 2018 has profoundly affected the way we collect and manage research data. Furthermore, the multiplication of open science plans, pledges, and funds – such as the French National Fund for Open Science, the San Francisco Declaration on Research Assessment (DORA) published in 2013,<sup>1</sup> the H2020 FAIR (Findable, Accessible, Interoperable, and Reusable) data principles and endeavors,<sup>2</sup> Plan S and cOAlition S,<sup>3</sup> the UNESCO 2021 Draft Recommendation on Open Science,<sup>4</sup> and several recent experiments in alternative publishing models – has shaken up the ecosystem of actors, norms, and practices in science. These

global trends engage various issues, from the need to respond to growing criticism of a commodification of science and the for-profit model of scientific publication by facilitating access to science as a global public good (Berkowitz & Delacour, 2020) to the need to address a rising prevalence of ethical scandals, scientific fraud and disavowals, and increasing need to assess data quality (Peng et al., 2021).

In this general context, open science or open research describes the transparent communication and sharing of research across the whole research lifecycle and research ecosystem from research stakeholders and research data to methods, software, tools, and publications (Willinsky, 2005). Recent evolutions toward open science vary across fields and disciplines. Indeed, in almost all disciplines, open publication is gaining ground and even becoming institutionalized, which translates into clearly identified and legitimate infrastructures and actors (like OpenEdition and Open Journal System), standards and seals (e.g., DOAJ), and a variety of models and initiatives that favor 'bibliodiversity'. Bibliodiversity is about diversity in the world of writing, reviewing, and publishing and sociomaterial practices (e.g., diamond open access journals, open peer

<sup>1</sup> <https://sfedora.org/>.

<sup>2</sup> <https://www.go-fair.org/fair-principles/> [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).

<sup>3</sup> <https://www.coalition-s.org/principes-du-plan-s/>.

<sup>4</sup> <https://en.unesco.org/science-sustainable-future/open-science/recommendation>.

\*Corresponding author: Héloïse Berkowitz, [heloise.berkowitz@univ-amu.fr](mailto:heloise.berkowitz@univ-amu.fr)

review, Peer Community in, and open monographies). In the very specific case of open research data, however, the situation is different.

The FAIR principles form the key framework guiding open research data (*The FAIR Data Principles – FORCE11*, 2014; Wilkinson et al., 2016). A few years later, in 2019, the Global Indigenous Data Alliance (GIDA) outlined a complementary framework, called the CARE principles (Collective benefit, Authority to control, Responsibility, Ethics) for Indigenous Data Governance. While the FAIR principles are generally widely known, applied and even required by funding agencies in several fields of research (Roche et al., 2015; Wilkinson et al., 2016), the CARE principles are less known and operationalized, despite their importance. Furthermore, there is a lack of clarity around what open research data, FAIR, and CARE principles really mean for social sciences. This issue has become even more important as incentives to open research data have met with concern and reluctance if not active resistance in the social sciences (Bonneville et al., 2021; Longley Arthur & Hearn, 2021).

We consider that the main arguments for opening data should be: to enhance science as a global public good via data patrimonialization, pooling, and synergies; verification and trust-building; broader engagement; add value to data production; facilitate and accelerate efforts to address the big challenges of our time, such as climate change; and other mission-oriented science. We also argue that while the ecosystem of actors, infrastructures, norms, standards, and principles is starting to take structure in France and abroad, there are several barriers to the process of opening data in social sciences that we need to address as a community, including: (1) a misperception of the motivations for opening data, that is, constructing a global public good rather than merely exercising control over researchers and their academic freedom, (2) a system based on competition, rankings, and the dominant process of 'starification' in research, (3) a lack of or unequal access to resources and capabilities that might further exacerbate inequalities among jobs, genders, communities, institutions, and countries, and (4) the potential risks inherent to open data and the specific constraints posed by the variety of types of social science data (e.g., interviews, field observations, pictures, economic datasets, and other quantitative databases or archives), all with various degrees of sensitivity and standardization challenges.

This paper sets out to inform a better understanding of what open research data would mean for social sciences. The paper is organized as follows. We start by providing a definition and an overview of the ecosystem of actors, infrastructures, and principles for open research data. We then briefly review the motivations for opening data, with a specific focus on data in social sciences. Next, using the emerging literature on open science and in particular open research data, we analyze

several of the factors that pose barriers to the process of opening data in social sciences. Finally, we investigate some practical questions for open research data, before presenting M@n@gement's open data policy and ending with some concluding thoughts.

## Overview of the open data ecosystem and research data lifecycle

### Definitions and ecosystem of actors

Before detailing the ecosystem, we first provide some clarifications and definitions concerning the specific terms related to open data, such as the different types of data access, metadata, data repository, and data management plan (DMP).

First, there are different types of research data access that vary in degree of openness: (1) fully open, with no barriers to access at all, (2) embargoed access, which means that external users cannot access datasets until the end of the embargo, (3) restricted access, with some barriers to access that external users can overcome under certain conditions, and (4) closed access, meaning totally closed access. There are also different categories of data, which affect and dictate their accessibility: that is, warm data, hot data, or cold data, depending on the frequency of accessibility. In parallel, there are different levels of data sensitivity, that is, data and information that has to be protected and/or restricted due to its sensitive nature (health, gender, beliefs, sexual orientation, etc.).

Second, metadata refer to key information about the data that facilitate their discoverability and therefore accessibility. Metadata structure and content are generally guided by a metadata standard and often vary across disciplines or repositories. There are various metadata standards, like DDI, COAR, and DCMI, that are all compiled in the Fair Sharing repository of standards (<https://fairsharing.org/>).

A third term commonly used is a data repository, which is where data are deposited. Published data need a permanent identifier like a DOI (Digital Object Identifier) that is usually provided by the repository. Different repositories serve different purposes, such as Zenodo and NAKALA for general data, ELIXIR for life sciences data, and CESSDA or DARIAH for social sciences and humanities data. Registries like Re3data (<https://www.re3data.org/>) provide an overview of existing repositories. Other types of meta-platforms, such as ISIDORE, harvest data produced in human and social sciences and then aggregate and process them to provide unified, enriched access. Similarly, OpenAIRE is a meta-organization and European Open Access infrastructure that harvests and aggregates open-access repositories, archives, and journals supporting open access. It also provides open data services like Amnesia, which is a tool for anonymizing data (<https://www.openaire.eu/amnesia-guide>). In France, Huma-Num is a large

research infrastructure dedicated to social sciences and humanities, which is also involved in DARIAH and served to create the NAKALA repositories, the Isidore platform, and a storage box, among other services.

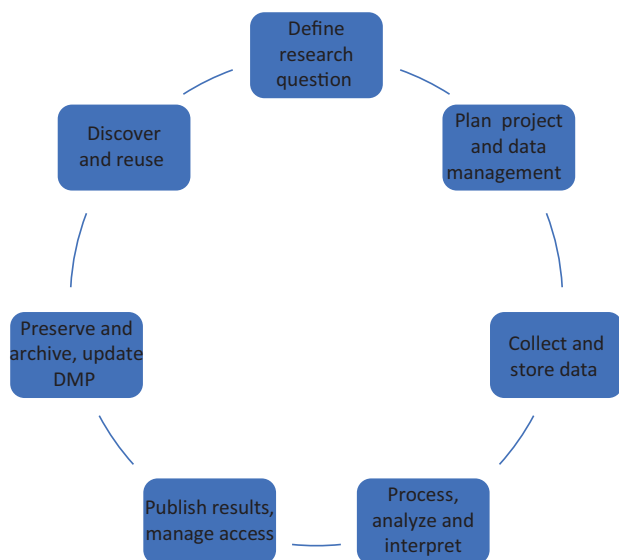
Fourth, most funding agencies require a DMP, which describes where, how, and when project data are to be collected, stored, and accessed. DMP Opidor, for instance, was purpose-developed to assist with DMPs. We discuss the DMP in more detail in section 4.

Complementary to these definitions, the open data ecosystem can be understood in terms of its stakeholders, that is, data producers (e.g., engineers, technical staff, and researchers), data publishers and data curators (e.g., standardization organizations, repositories, and other platforms), funders and sponsors who generally impose accessibility requirements, and data users (e.g., scholars, decisionmakers, and citizens) who make use of the openly accessible data.

### Research data lifecycle and workflows

The objectives tied to data reuse and patrimonialization mean that research datasets actually have a longer lifespan than the research projects themselves. The research data lifecycle is illustrated in Figure 1.

Based on a given defined research question, a DMP involves identifying which kinds of data are needed and are to be collected and then managed. At this stage, a first version of a DMP is usually created, typically as part of an application for funding. This draft DMP covers some key issues, such as where data will be stored and whether and how it will be accessed.



**Figure 1.** Research data lifecycle (adapted from the Princeton Research Lifecycle Guide (<https://researchdata.princeton.edu/research-lifecycle-guide/research-lifecycle-guide/>))

Then, during the fieldwork or experimental stage, researchers will collect data, conduct surveys or interviews, consult archives and other data formats, and describe and record the data collection methods.

Once data have been collected, it must be processed, cleaned, combined, selected, or removed. Data processing also needs to be documented to ensure quality control and replication; only then can the data be analyzed and interpreted.

For publication purposes, it might be necessary to give access to both raw and processed data, to data visualizations, and data documentation, or make it directly available to reviewers via data platforms (with closed or open access rights). At this stage, there may be data validation, either through the normal reviewing process or through a purpose-dedicated data validation process.

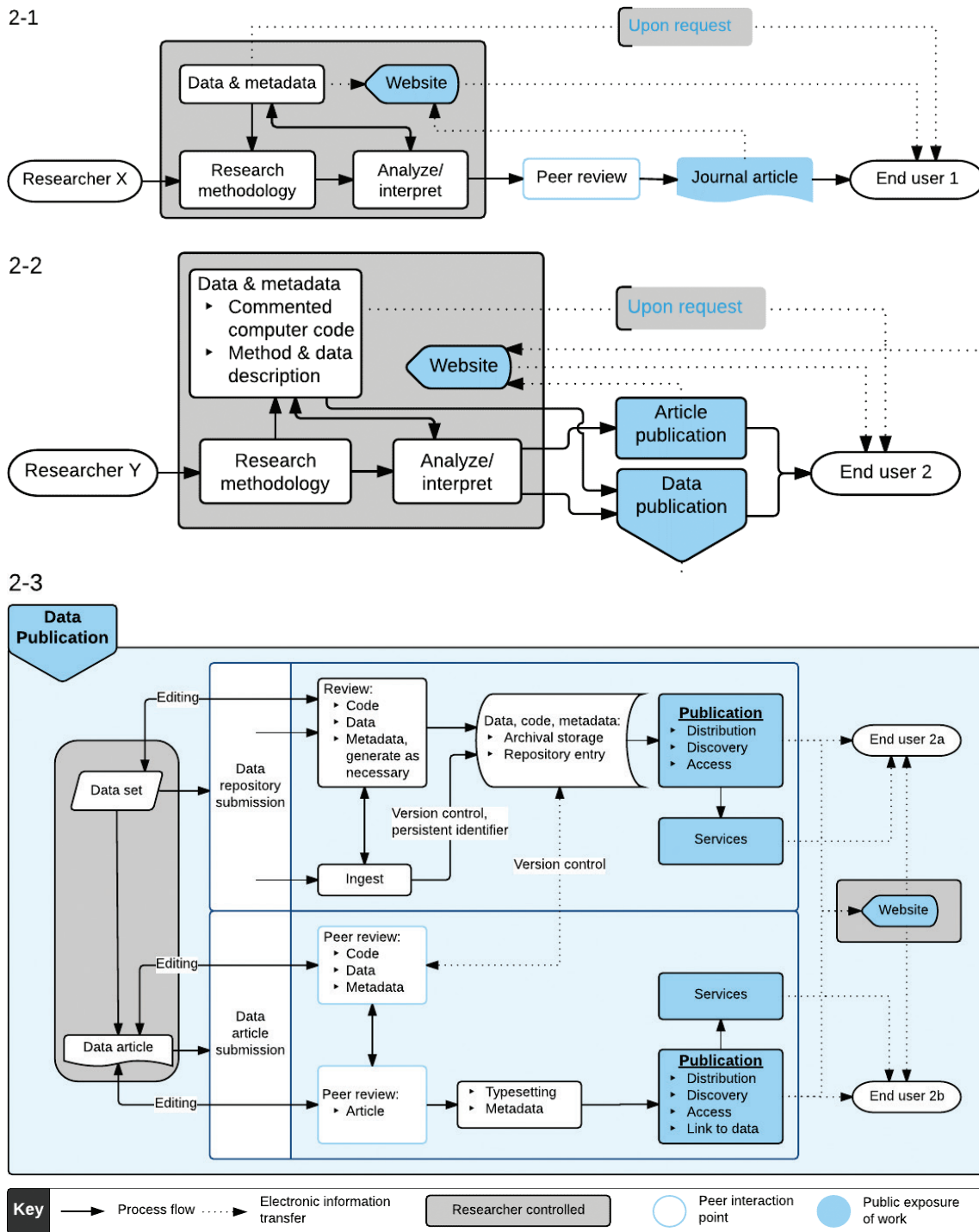
Toward the end of the project, data sharing may require proper preservation and archive practices to ensure that the data remain accessible and reusable long-term. Preparation for data preservation and archival might follow a different workflow to data collection or processing. Among other elements, it will involve description of metadata, attribution of a permanent identifier (enabling citation, such as a DOI number), licensing for reuses, storage in a data repository, and management of data access (depending on data sensitivity).

There may be some variations along this schematic lifecycle as and where dictated by data sensitivity and accessibility imperatives. To further illustrate this issue, we borrow Austin et al.'s (2015) detailed description of data publishing components and workflows (Figure 2), as the figures are easily readable and helpfully show the growing complexity of the workflows, especially in the case of data papers, that is, papers that describe and accompany datasets.

Now that we have presented the research-data lifecycle and workflows in the context of open science, and we turn to the key guiding frameworks and principles that have emerged to structure the practice of open research data.

### Guiding frameworks and principles

In 2014, FORCE11 published a Joint Declaration of Data Citation (JDDC). FORCE11 is a community of librarians, researchers, publishers, and various institutions dedicated to changing practices toward more open, transparent, and collaborative research creation and sharing. This declaration constituted one of the first acknowledgements of data as a significant and valuable research output per se. It further sought to foster best practices around data citation and reuse by outlining a set of guiding principles for data citation (data citation synthesis group: joint declaration of data citation principles, 2014) (Table 1).



**Figure 2.** Research data publication workflows reproduced from Austin et al. (2015, p. 16): Traditional article publication workflow (Fig. 2-1), reproducible research workflow (Fig. 2-2), and open data publication workflow (Fig. 2-3)

On this basis, a series of principles have emerged that frame and guide the functioning of open research data. These 'FAIR' principles were again developed by

FORCE11, in 2016. The FAIR principles dictate that open research data should meet the following requirements (Table 2).

**Table 1.** The joint declaration of data citation principles**1. Importance**

Data should be considered legitimate, citable products of research. Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications.

**2. Credit and attribution**

Data citations should facilitate giving scholarly credit and normative and legal attribution to all contributors to the data, recognizing that a single style or mechanism of attribution may not be applicable to all data.

**3. Evidence**

In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited.

**4. Unique identification**

A data citation should include a persistent method for identification that is machine-actionable, globally unique, and widely used by a community.

**5. Access**

Data citations should facilitate access to the data themselves and to such associated metadata, documentation, code, and other materials as are necessary for both humans and machines to make informed use of the referenced data.

**6. Persistence**

Unique identifiers and metadata describing the data and its disposition should persist — even beyond the lifespan of the data they describe.

**7. Specificity and verifiability**

Data citations should facilitate identification of, access to, and verification of the specific data that support a claim. Citations or citation metadata should include information about provenance and fixity sufficient to facilitate, verifying that the specific timeslice, version, and/or granular portion of data retrieved subsequently is the same as was originally cited.

**8. Interoperability and flexibility**

Data citation methods should be sufficiently flexible to accommodate the variant practices among communities but should not differ so much that they compromise interoperability of data citation practices across communities.

Source: <https://force11.org/info/joint-declaration-of-data-citation-principles-final/>

Some years later, in 2019, the GIDA outlined a complementary framework to the FAIR principles, called the CARE principles for Indigenous Data Governance. These principles were established to advance data rights in the context of the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP) and to ensure that data production, collection, storage, diffusion, and citation benefit indigenous peoples (Carroll et al., 2021). The CARE principles are defined as in Table 3.

Operationalizing these principles is crucial. The Integrated Data Infrastructure (IDI) in New Zealand tackles this issue through a protocol developed to ensure that data comply with both FAIR and CARE principles

**Table 2.** FAIR principles**To be Findable**

- F1. (Meta)data are assigned globally unique and eternally persistent identifiers.
- F2. Data are described with rich metadata.
- F3. Metadata specify the data identifier.
- F4. (Meta)data are registered or indexed in a searchable resource.

**To be Accessible**

- A1. (Meta)data are retrievable by their identifier using a standardized communication protocol.
  - A1.1. The protocol is open, free, and universally implementable.
  - A1.2. The protocol allows for an authentication and authorization procedure where necessary.
- A2. Metadata should be accessible, even when the data are no longer available.

**To be Interoperable**

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow the FAIR principles.
- I3. (Meta)data include qualified references to other (meta)data.

**To be Reusable**

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes.
  - R1.1. (Meta)data are released with a clear and accessible data usage license.
  - R1.2. (Meta)data are associated with detailed provenance.
  - R1.3. (Meta)data meet domain-relevant community standards.

Source: <https://www.go-fair.org/fair-principles/>.  
FAIR, Findable, Accessible, Interoperable, and Reusable.

(Carroll et al., 2021). This protocol, which we have reproduced here in Figure 3, can serve as a source of inspiration for all.

These principles are known and agreed upon but are often not directly applied by the whole researcher community. It is generally funders who specify that the research data are to be 'open', but without giving any further guidance. For instance, the French National Research Agency (ANR) requires a DMP to be provided 6 months after beginning a project. The FAIR principles are compulsory in some settings, such as for securing European fundings. However, few social sciences journals even have an open data policy, despite the fact that they increasingly demand datasets to be provided for reviewing processes and to enable readers to consult online datasets for accepted papers. This emerging trend can be expected to become generalized to most journals and may even, in some cases, become a requirement. Given this context, what are the actual motivations and challenges to opening data in social sciences?

**Table 3.** CARE principles**Collective benefit**

Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data.

**C1. For inclusive development and innovation**

Governments and institutions must actively support the use and reuse of data by Indigenous nations and communities by facilitating the establishment of the foundations for Indigenous innovation, value generation, and the promotion of local self-determined development processes.

**C2. For improved governance and citizen engagement**

Data enrich the planning, implementation, and evaluation processes that support the service and policy needs of Indigenous communities. Data also enable better engagement between citizens, institutions, and governments to improve decision-making. Ethical use of open data has the capacity to improve transparency and decision-making by providing Indigenous nations and communities with a better understanding of their peoples, territories, and resources. It similarly can provide greater insight into third-party policies and programs affecting Indigenous Peoples.

**C3. For equitable outcomes**

Indigenous data are grounded in community values, which extend to society at large. Any value created from Indigenous data should benefit Indigenous communities in an equitable manner and contribute to Indigenous aspirations for wellbeing.

**Authority to control**

Indigenous Peoples' rights and interests in Indigenous data must be recognized and their authority to control such data be empowered. Indigenous data governance enables Indigenous Peoples and governing bodies to determine how Indigenous Peoples as well as Indigenous lands, territories, resources, knowledge, and geographical indicators are represented and identified within data.

**A1. Recognizing rights and interests**

Indigenous Peoples have rights and interests in both Indigenous knowledge and Indigenous data. Indigenous Peoples have collective and individual rights to free, prior, and informed consent in the collection and use of such data, including the development of data policies and protocols for collection.

**A2. Data for governance**

Indigenous Peoples have the right to data that are relevant to their world views and empower self-determination and effective self-governance. Indigenous data must be made available and accessible to Indigenous nations and communities in order to support Indigenous governance.

**A3. Governance of data**

Indigenous Peoples have the right to develop cultural governance protocols for Indigenous data and be active leaders in the stewardship of, and access to, Indigenous data, especially in the context of Indigenous knowledge.

**Responsibility**

Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples' self-determination and collective benefit. Accountability requires meaningful and openly available evidence of these efforts and the benefits accruing to Indigenous Peoples.

**R1. For positive relationships**

Indigenous data use is unviable unless linked to relationships built on respect, reciprocity, trust, and mutual understanding, as defined by the Indigenous Peoples to whom those data relate. Those working with Indigenous data are responsible for ensuring that the creation, interpretation, and use of those data uphold, or are respectful of, the dignity of Indigenous nations and communities.

**R2. For expanding capability and capacity**

Use of Indigenous data invokes a reciprocal responsibility to enhance data literacy within Indigenous communities and to support the development of an Indigenous data workforce and digital infrastructure to enable the creation, collection, management, security, governance, and application of data.

**R3. For Indigenous languages and worldviews**

Resources must be provided to generate data grounded in the languages, worldviews, and lived experiences (including values and principles) of Indigenous Peoples.

**Ethics**

Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem.

**E1. For minimizing harm and maximizing benefit**

Ethical data are data that do not stigmatize or portray Indigenous Peoples, cultures, or knowledge in terms of deficit. Ethical data are collected and used in ways that align with Indigenous ethical frameworks and with rights affirmed in UNDRIP. Assessing ethical benefits and harms should be done from the perspective of the Indigenous Peoples, nations, or communities to whom the data relate.

**E2. For justice**

Ethical processes address imbalances in power, resources, and how these affect the expression of Indigenous rights and human rights. Ethical processes must include representation from relevant Indigenous communities.

**E3. For future use**

Data governance should take into account the potential future use and future harm based on ethical frameworks grounded in the values and principles of the relevant Indigenous community. Metadata should acknowledge the provenance and purpose and any limitations or obligations in secondary use, inclusive of issues of consent.

Source: <https://www.gida-global.org/>.

UNDRIP, United Nations Declaration on the Rights of Indigenous Peoples.



Figure 3. The Ngā Tikanga Paihere guidelines (Stats NZ, 2020, p. 10)

**Motivations for open research data**

The rationale for opening data has been detailed and defended first and foremost by public, national, transnational, or international governmental or nongovernmental organizations like the European Commission, the French National Research Council (CNRS), FORCE I I, or the OECD, based on a push for greater transparency, accessibility, and accountability.

The main arguments and reasons for opening research data from publicly funded research include: to enable more accurate data verification; reduce risks of duplication and costs in data collection and production; facilitate data pooling and, therefore, reuse; foster broader engagement not only from other researchers but also from citizens, governments, and organizations; create trust in science (OECD 2016); facilitate multidisciplinary research on complex societal challenges (particularly in climate-dependent social sciences; Gauquelin et al.,

2017); lend value to what is sometimes considered invisible work (i.e., data collection and production); and, ultimately, ensure patrimonialization, that is, safe storage and preservation for future uses, as a cultural heritage patrimoine. Patrimonialization may be one of the least known yet most important reasons for the development of open research data. Indeed, considering current datasets as future archives opens promising avenues for future research. Research data collected and produced today – especially when publicly funded – hold a heritage value and, as such, must be preserved and made available under certain conditions.

Other arguments can also be advanced. OpenAIRE, the open access infrastructure for research in Europe, provides a slightly different list of benefits of open research data for a variety of stakeholders.<sup>5</sup> OpenAIRE argues that benefits of

<sup>5</sup> Source : <https://www.openaire.eu/what-is-open-research-data>.



open research data are as follows: for researchers, greater discoverability, increased efficiency, new collaborations, and more funding; for funders, increased visibility and reuse of funded research, greater impact, and return on investment; for the general public, self-empowerment, increased transparency, and greater engagement in science and research; for organizations and NGOs, better access to research, information-sharing, and more effective advocacy and lobbying; for national governments, the promise of data-driven decision-making, reduced government costs, and more effective and efficient government services.

However, these arguments are based on two assumptions. First, science and research data should help achieve better performance and efficiency, rather than ensuring social and ecological justice and serving the common good. This more generally raises the issue of commercial use of open research data. Second, this assumes that all listed stakeholders, including governments and organizations, actually have the capacity to use and reuse open data. However, data use and data-informed decision-making require specific capabilities and infrastructures that are not yet all available. Simply opening up data does not guarantee that government services, for example, will be more efficient.

From our perspective as researchers and co-editors in chief of M@n@gement, the most important arguments in favor of opening data in social sciences are patrimonialization, data pooling and synergies, verification and trust-building, broader engagement, data production mobilization, and facilitation and acceleration of efforts to address the big challenges of our time, such as climate change and other mission-oriented science, because all of these benefits help enhance science as a global public good. We now turn to challenges and barriers.

### Challenges and barriers to opening data in social sciences

This section is not about the 'open data excuse bingo' or the 'concerns about opening up data, and responses which have proved effective', or 'how to make friends and get them to give you their data', as argued by Gutteridge & Dutton in their list of excuses.<sup>6</sup> This fairly exhaustive list provides an interesting overview of common concerns about open data and their solutions, including risks of terrorist uses, misinterpretation of data, data protection laws, authorization and ownership, or (poor) data quality. Here, we focus on what we consider as the most important challenges and barriers to opening data in social sciences.

<sup>6</sup> See here for their collaborative Google document [https://docs.google.com/document/d/1nDtHpnIDTY\\_G32EMJniXaOGBufjHCck4VC9WGOif7jK4/edit#](https://docs.google.com/document/d/1nDtHpnIDTY_G32EMJniXaOGBufjHCck4VC9WGOif7jK4/edit#).

### Misperception of motivations for open research data

One main obstacle to opening research data in science in general and in social sciences in particular has to do with the reasons and rationales for open data. Social scientists often either ignore or misperceive motivations for opening data. We – including ourselves as the authors of this piece – often view opening data as a way for authorities to exercise control over academic freedom and over researchers, or simply as an additional workload burden. But other reasons detailed above, like patrimonialization of data, data pooling and synergies, verification and trust-building, broader engagement, and facilitation of mission-oriented science, are much more important rationales that merit consideration. Another reason open data might be conceived as a general threat is because we often miss the fact that open data are actually about being 'as open as possible, and as restricted as necessary', which means a lot of freedom and variants.

### Dominance of competition and the starification process

Science in general and social sciences and management and organization studies in particular oscillate between a process of revering top-ranking researchers as stars and calling for collaborative, horizontal relations. Evidence shows that in science, flat teams are more innovative and have more long-term impact than hierarchical teams (Xu et al., 2022). Furthermore, some fields in particular are heavily organized by the effect of rankings (of schools, journals, etc.), which aggravates the star-system (Osterloh & Frey, 2015). Management science even has the added specificity of a double system of public research in public institutions and private research in business schools whose financial bonuses are indexed to publication output. Open research data, on the other hand, are highly collaborative, often co-constructed to facilitate reusability, which can prove contradictory with a competitive ranking-driven paradigm. This tension between two paradigms connects with the strong lock-in of academic assessment, which still overwhelmingly emphasizes individual publications in ranked or high-impact-factor journals, despite positive ongoing change as promoted by the CNRS or Utrecht University, which favors open science values (see Woolston, 2021).

This tension between paradigms also more generally relates to an issue of attribution of intellectual property or recognition of data production (Gauquelin et al., 2017). Some researchers might simply not be willing to share their data, which they consider their property. Furthermore, some might even claim that data cannot be shared openly

and at no cost for the competition, regardless of whether other potential users are or are not viewed as competitors. In other cases, those who effectively produced data (such as engineers or research assistants) are rare if ever acknowledged or considered as coauthors, which raises deontological issues. But opening data might further change the balance of power, as researchers might get recognition for the data produced, whereas the engineers and other technical support colleagues may become (even more) invisible (Gauquelin et al., 2017; Millerand & Bowker, 2008).

### **Costs and lack of resources and capabilities**

A further challenge posed by opening research concerns the costs involved and the lack of resources and capabilities to fulfill new requirements. Indeed, in social sciences, datasets are often produced by a single researcher or a small team of researchers (Burgelman et al., 2019). Data collection in social sciences is highly time-consuming. Managing open data requires additional time and effort. A decoupling could appear between researchers' requirements and perceptions of the job (Longley Arthur & Hearn, 2021). Furthermore, while some funding agencies readily cover some of the costs connected to preparing data, this is not widespread practice and is not generalized in all types of funding instruments, despite the increasing open data requirements. This situation may create or aggravate inequalities among researchers and institutions, between those who have the resources or institutional support and those who lack these capabilities.

Publishers are increasingly requesting data availability statements to indicate where and how datasets are accessible, but most institutions or repository staff still provide only limited support (Longley Arthur & Hearn, 2021). In that respect, Austin et al. (2015) regret that scholars are asked to adapt to emerging infrastructures instead of being provided platforms constructed around the way they already produce data.

All these challenges are related to the lack of resources and capabilities available to a researcher or an open access journal, as the costs required for open research data might further widen inequalities among researchers, genders, communities, journals, institutions, and countries.

### **Constraints and risks specific to social sciences**

Among the myriad challenges faced in the social sciences, the research communities are not used to sharing pre-print online and sharing data in general, which makes it harder to change practices (Longley Arthur & Hearn, 2021).

More importantly, one specificity of social sciences lies in the wide range of types and formats of data collected, which

can range from quantitative to qualitative datasets, supported by interviews, logbooks, personal notes, confidential files, or archives from organizations and institutions under observation, to name but a few. Mixed methods are also gaining ground. This raises important issues around data standardization practices, costs of working with different formats, and property ownership. The property ownership issue, for instance, assumes that institutions that self-publish archives will more widely adopt an open data policy and share their data according to FAIR and CARE principles. Creative Commons Attribution licenses could offer an opportunity to support this wider access to open data and facilitate social science work.

Another particularity of social sciences is data anonymity and de-identification, which is often confused with confidentiality (Borgerud & Borglund, 2020). Replacing identity markers by an identifier that has no meaning to anyone has long been traditional practice in research, but now, in the era of big data, it offers no real assurance that a data subject's identity is protected. In parallel, if sensitive personal data, such as personal notes or logbooks, cannot be protected through anonymity, then it becomes necessary to restrict access to the data to ensure its integrity. Password protection and signing ethical agreements would mean that only authorized persons could access, use, or reuse these data.

On the confidentiality front, there are already extensive sets of rules and recommendations developed, by the OECD (2016), for instance. Confidentiality and privacy protection are a complex issue, with both legal and ethical aspects that should be addressed in the DMP (see below). Clarifying how privacy will be ensured when using human data is, of course, vitally important. The concept of consent is also at the core of this ethical approach. Regarding open data, consent concerns not only how individuals are informed of the way the data will be collected and used, but also its potential reuse by other researchers. Furthermore, in social sciences, protection extends not just to data but also to interviewees (Bonneville et al., 2021). One major risk for research is losing access to fieldwork, either because interviewees are less willing to participate in open research data projects or because conducting such projects becomes too expensive under open access conditions.

### **Operational conditions for opening research data**

Against this backdrop, we now turn to explore the operational conditions and concrete questions that we should be asking when opening research data. We focus on researchers on the one hand and independent journals on the other hand.

Practice 'CARE' in data collection	Engage 'CARE' in data stewardship	Implement 'CARE' in data community	Use 'FAIR' with 'CARE' in data applications
Define cultural metadata Record provenance in metadata	Use appropriate governance models Make data 'FAIR'	Indigenous ethics inform access Use tools for transparency, integrity and provenance	Fairness, Accountability, Transparency Assess equity

**Figure 4.** Implementation of CARE principles (Carroll et al., 2021, p. 4)

### For researchers

Beyond questions related to the data lifecycle and workflows mentioned earlier, there are at least five key questions to ask when considering opening data.<sup>7</sup> First, can your data be shared and under what (ethical) conditions? Second, how do you prepare your data for sharing? Third, how do you find and choose a repository? Fourth, what is a data availability statement and what should you detail in your paper? Fifth, how do you link your datasets (and others) to your paper? These questions are generally addressed in DMPs, but broadly speaking, they should be tackled collectively with all those who have been involved in collecting, producing, and processing the data, that is, data stewardship. This means co-authorship of data should also be addressed.

First, one should keep in mind that open research data mean as 'open as possible, and as restricted as necessary'. Addressing the conditions for data sharing requires thinking about the way data were or will be collected, and the various limits and barriers to making data available (e.g., whether it is sensitive or not). Integrating FAIR and CARE principles requires additional reflection regarding relations to community, as shown in Figure 4 (Carroll et al., 2021).

Second, the way data are prepared for sharing will vary based on the type of data being shared and the rules and expectations of repositories and disciplines. The COAR (Controlled Vocabularies for Repositories) provide a standardized description of data concepts or types.<sup>8</sup> Social sciences use a great variety of data, from images to text corpuses, social media data, surveys, and large-scale datasets, among others. Qualitative data that can be openly made available on repositories may include everything from structured, semi-structured, or unstructured interviews to focus groups; oral narratives in the form of audio or video recordings; transcripts, notes, and summaries; field notes (including from participant observation or ethnography); maps, satellite imagery, or geographic data; official documents, files, reports,

minutes of meetings, transcripts of public speeches, autobiographies, memoirs, travel logs, diaries, brochures, posters, flyers; personal documents like letters, diaries, correspondence papers; radio broadcasts whether audio or transcripts; TV programs; photographs; and more. In general, there are five specific 'Dublin-Core'-inspired standard fields of meta-data: title, author, date, type, and license. Other Dublin-Core fields can be added depending on the data types: contributor, coverage, creator, description, language, publisher, relation, rights, source, subject, and more.

Third, the choice of a repository depends on the discipline and type of data but, more importantly, on several minimum criteria that ensure trustworthiness and FAIR principles (Science Europe, 2021). Trustworthy repositories should provide (1) persistent and unique identifiers that enable searching, identifying, citing, and retrieving data; (2) organized, findable, standardized, and machine-retrievable meta-data; (3) data access and usage licenses with well-specified access conditions, ensuring data authenticity and integrity, and respectful of confidentiality and data subjects' and creators' rights; and (4) preservation mechanisms that ensure persistence, while being transparent about their mission, models, and sustainability, all in compliance with FAIR data sharing and access principles. Some repositories, like NAKALA or the Qualitative Data Repository (QDR), can store most if not all formats of qualitative data.

Fourth, including a data availability statement is increasingly a requirement from all journals, even when no data were used. This statement should be added to manuscripts before submission. For cases with no data, the following statement can be used: 'No data is associated with this article'. For articles where all associated data are described and presented within the manuscript, the statement can say that 'All data supporting the results is available as part of the article, and no additional source data is required'. Then, when data are hosted in a repository, statements should generally mention the title and DOI (for repositories like Zenodo or QDR) or give the title and DOI and embed code for interactive reanalysis tools if mixed data and code are deposited (Code Ocean). If data cannot be shared, meta-data can still be deposited and mentioned in the data availability statement. Exceptions include

<sup>7</sup> See <https://www.ouvrirlscience.fr/partager-les-donnees-liees-aux-publications-scientifiques-guide-pour-les-chercheurs/> and <https://open-research-europe.ec.europa.eu/for-authors/data-guidelines>.

<sup>8</sup> [https://vocabularies.coar-repositories.org/resource\\_types/](https://vocabularies.coar-repositories.org/resource_types/).

confidentiality, trade secrets, security rules, Union competitive interests, or intellectual property rights, or if it is impossible to adequately de-identify human data, in which case the data availability statement gives details.

Fifth and last, once a manuscript has been accepted and published, it is recommended to update the DMP and repository information with a link to the final article, and vice versa. This reciprocal connection is important for visibility and accessibility.

All these elements are addressed in the DMP. There are many resources available online to provide guidance on designing DMPs. Here, we briefly review the key sections of a DMP in relation to the questions set out above and based on the Science Europe (2021) extended guide (see Table 4). A DMP generally starts with administrative information, such as name of applicant, project number, funding program, and version of DMP. The next important section describes the data and the collection or reuse of existing data. It addresses how data will be collected or produced, what kind of data, formats, and volumes will be used, and whether and which data quality control

**Table 4.** Data management plans: Issues for researchers (from Science Europe, 2021)

Section	Information and issues addressed
General information	Name of applicant, project number, funding program and number, version of DMP, etc.
Data description and collection or reuse of existing data	How data will be collected or produced, what kind of data, formats, and volumes, whether and which data quality control measures will be used, etc.
Documentation and data quality	Metadata, standards, methodology describing data collection, measures for quality control, etc.
Storage issues and backup solutions (during research)	Storage, backup, data security, protection of sensitive data, etc.
Legal and ethical requirements and codes of conduct	How personal data are processed in compliance with the GDPR and other data privacy and protection laws, how intellectual property rights and ownership are managed, how ethical issues and codes of conduct arise and are managed in the research, etc.
Data sharing and long-term preservation mechanisms	Possible restrictions or reasons for embargo, tools for data use and reuse, attribution of unique and persistent identifiers, etc.
Data management responsibilities and resources	Who is responsible for data stewardship, what resources will be dedicated to it, etc.

DMP, Data Management Plan; GDPR, General Data Protection Regulation.

measures will be used. Then, the DMP details documentation and data quality, that is, meta-data provided, standards used, methodology describing data collection, and measures to ensure quality control. The next section deals with storage issues and backup solutions employed during the research process itself, including data security and protection of sensitive data. Then, a DMP must address legal and ethical requirements and codes of conduct, detailing how personal data are processed in compliance with GDPR and other data privacy and protection laws, how intellectual property rights and ownership are managed, which ethical issues and codes of conduct arise in the research, and how they are managed. Next, the DMP explains data sharing and long-term preservation mechanisms, possible restrictions or reasons for embargo, tools for data use and reuse, and the attribution of unique and persistent identifiers. The last section generally deals with data management responsibilities and resources, that is, who is responsible for data stewardship and what resources will be dedicated to it.

All these questions to ask and aspects of data management to plan might seem overwhelming at first, but forward-planning, reflexivity on data needs, time, training, and support from parent institutions all help to make sharing open data a much more achievable and rewarding task.

### For journals

We now turn to practical conditions and limits for journals. Why this focus on journals, and not, say, on academic societies, research labs or universities that clearly play an important role in this process? Journals are at the forefront of these issues, and here, at *M@n@gement*, we asked ourselves several key questions when we started developing our open data policy. Some of the questions that we asked ourselves were: What kind of guidelines can a journal develop? Should open datasets be reviewed and by whom? Who should bear the cost of reviewing data? What should we do with data papers? At this stage, some of these questions remain unresolved.

Our self-interrogation is driven by one major issue: what are the roles and responsibilities of a journal like ours in the process of opening research data? At *M@n@gement*, we acknowledge the importance of exploring ways to improve open science, which includes facilitating open data. However, as an independent, in-house, diamond open access journal, one of our first concerns is the human and financial costs of implementing and advocating an open data policy. We can define guidelines and make them an incentive, but we cannot develop our own infrastructure for instance, and we believe this is not the responsibility of journals like ours.

Looking at existing journal policies, few already have an open data policy. In economics, for instance, the *American*

*Economic Review* has been a standard-setter, and its data policy has been adopted by other journals, including the *Journal of Political Economy*. These economics journals make sharing data compulsory upon submission, and they use Harvard Dataverse, a repository for research data where journals can create their own space. In the case of *M@n@gement*, we believe that using public non-university-based repositories (like NAKALA) would allow more bibliodiversity in data sources and avoid having data stored in North America (which raises specific security and ownership issues). In the medium-to-long term, the sustainability of repositories and competition among repositories might become an issue, which makes it all the more important to choose a renowned, institutionalized repository.

Table 5 synthesizes the issues and questions to address when defining a journal open data policy (based on Feret et al., 2021; Hrynaszkiewicz et al., 2020). The objective of this kind of policy and the topics it should address is to identify which data are concerned, at which stage of the process, and which standard and metadata protocols should be used (see also Marlet et al., 2022).

In line with the availability of platforms and repositories, at the journal level, there are tensions between open access journals and open research data systems: journals with ample resources from commercial editors might find it easier to adapt to these changes, which are already ongoing at some

journals. However, diamond open access journals from independent or nonprofit organizations might find it harder if they are expected to deal with open research data on their own. Journals, especially independent ones, cannot be expected to provide the infrastructure for handling or even reviewing open data. Data stewardship is a specific profession that is very different from academic publishing. However, in relation to these general questions, one of the arguments for open science is to move away from profit-oriented publication. So, this raises a question: should we allow to make any kind of profit from research data or from publishing research? Or should we at least seek to ensure that commercial uses first and foremost benefit public organizations, universities, and citizens?

## The M@n@gement open data policy

*M@n@gement* has decided to encourage authors to make all data associated with their submissions openly available, whenever possible and with all necessary restriction conditions, in accordance with the FAIR and CARE principles. Our policy is only an incentive, but we hope to raise awareness of 'open' issues and the need to integrate FAIR and CARE principles.

To develop the *M@n@gement* open data policy, we used the Science Europe (2021) extended guide and borrowed other journals' policies for inspiration (for instance, *Designs for learnings*) as well as guidelines developed by the Research Data Alliance data policy standardization and implementation interest group (Fabre & Gouzi, 2020; Hrynaszkiewicz et al., 2020) and the Research Data College of the French Committee for Open Science (Feret et al., 2021).

We consider this document to be a preliminary draft and subject to change, both in response to institutional evolutions, to potential consolidations in the open science landscape, and to the evolutions of our own field. We mostly seek to open the debate with our community about the needs, challenges, and perspectives for open research data. We generally follow the French Committee for Open Science's principle of 'as open as possible, as restricted as necessary'.

## Definition of open research data

This policy applies to research data that would be necessary to check the results presented in the journal's publications. 'Research data' include data produced by the authors as well as data from other sources that are analyzed by the authors in their study. These data can be presented in various forms, such as image, video, text, code, or statistical table but must be produced and shared in compliance with FAIR and CARE principles (<https://www.rd-alliance.org/implementing-care-principles-care-full-process>). Research

**Table 5.** Roadmap for defining a data policy for journals

General topic	Issues and questions to address
Definition of research data and exceptions	What are the access and embargo conditions? Which data are concerned by the policy? At which stage of the process? Submission? Acceptance?
Data and metadata standards and formats	Which standards should be used and what are the metadata protocols?
Data access, hosting, and publishing	What are the protocols and guidelines for depositing data and choosing a repository?
Data availability procedures	What are the data availability procedures (timeframe, stages, etc.)? Are data peer-reviewed and by whom? Is there a data availability statement?
Data accessibility	How will data be connected to the publication? Are permanent identifiers used?
Noncompliance	What happens when authors do not comply with the journal's data policy?
Support for authors, reviewers, and editors	Does the journal or its academic association offer some form of support?

data that are not necessary for findings reported in publications are not covered by this policy.

This policy will be limited by the legitimate exceptions regulated by law, for example, with regard to professional confidentiality, trade secret, personal data, or content protected by copyright. *M@n@gement* does not review and publish data papers, but we encourage authors to add data papers to their datasets as presentations in order to clarify data procedures, conditions, limits, or reuses.

### **Data and metadata standards and formats**

*M@n@gement* encourages authors to use open and standard formats. Descriptive metadata must be structured using recognized standards, at least Dublin Core. Standards used by researchers can be either discipline-specific or more generic ([https://en.wikipedia.org/wiki/Dublin\\_Core](https://en.wikipedia.org/wiki/Dublin_Core)).

*M@n@gement* advises authors to use 'controlled' (or reference) vocabularies, whether discipline-specific or more generic, to express metadata (e.g., to reference an author, see <https://orcid.org>; to reference a place, see <https://www.geonames.org>; for data concepts, see *Controlled Vocabularies for Repositories*). We also advise using data file formats that comply with CINES recommendations for long-term preservation (<https://facile.cines.fr>; in French).

### **Data access and hosting**

The data that contributed to writing the paper must be deposited in a data repository that will guarantee secure storage and access to the data, in particular through the attribution of a permanent identifier, such as a DOI. We advise authors to avoid using private repositories whose roadmap is not transparent in terms of economic model, governance, or sustainability (e.g., Figshare).

*M@n@gement* recommends the use of trustworthy, FAIR-compliant repositories, whether they are generalist (e.g., Zenodo), institutional (e.g., Data INRAE or university platforms), or discipline-specific (e.g., QDR or NAKALA for social sciences and humanities). In all cases, authors should check that the chosen repository meets minimum quality criteria (see <https://zenodo.org/record/4915862>).

We, therefore, encourage authors to contact their institution's data management, sharing, and stewardship support services for help and advice on good DMP design and development practices.

### **Data availability procedure**

The data availability procedure follows three steps.

Submission phase

Authors are encouraged to transmit the data along with their submission whenever possible. This can be done either within the article or in appendix (dataset option on our platform) but preferably through a restricted or controlled-access repository. All such data must remain anonymized and adhere to our general recommendations. This includes anonymization of the research project itself, principal investigator, and research participants.

Peer review phase

If editors and reviewers deem it necessary, the authors should make the data supporting results reported in their contribution available for reviewers. Papers will be rejected if authors refuse to provide data when asked.

Acceptance phase

To the extent possible, data should be made available without embargo or with the shortest embargo period when the paper is accepted. The terms for sharing must allow reuse, with an explicit link between the data and the publication they support, under normal conditions (in compliance with guidance on personal data and protection of interviewees).

While *M@n@gement* encourages authors to share data under open licenses that allow free reuse, authors must be careful to use the licenses recommended by the repository where the datasets were deposited.

By publishing in *M@n@gement*, authors commit to make the data and/or metadata publicly available for at least 5 years after their contribution has been published, either through a platform or by individual provision if the data cannot be freely shared.

Alternatives to open-access sharing of personal or sensitive data are:

- Anonymization or pseudonymization of the data before open access release;
- Data available on request for research purposes only;
- Availability of the metadata only, which should be the minimal objective for all authors submitting to *M@n@gement*.

### **Data accessibility statement**

Authors are expected to cite the datasets underpinning their publication in a specific data accessibility statement, which is to describe the available data and how to access it and provide a permanent link to the data.

The statement may include one or more of the following options:

- The datasets generated during and/or analyzed during the current study are available in the [NAME] repository; [DOI].

- The datasets generated during and/or analyzed during the current study are not available in open access due to [specify reasons] but are available from the author on reasonable request.
- Data sharing does not apply to this article because no datasets were generated or analyzed during the current study.
- The datasets on which the current study is based were not generated by the authors. They are available online: Creator (year of publication), Title, Version. [Repository Name] [DOI].

## Conclusion

Recent international trends highlight efforts to 'open' science across its whole ecosystem and lifecycle – from capturing research data through publishing results. However, opening research data raises specific issues and concerns for the field (Bonneville et al., 2021; Gauquelin et al., 2017). Here, we reported a first overview of the existing ecosystem of actors, infrastructures, standards, and guiding principles and the reasons for opening data. We also showed several barriers to opening research data: (1) a misperception of the motivations for opening data (i.e., patrimonialization, data pooling and potential synergies, trust-building, and broader engagement, rather than merely exercising control over researchers and their academic freedom and researchers), (2) a system based on competition and the dominant process of 'starification' in research that encourages researchers to focus on publications and retain ownership of 'their' data, (3) a lack of resources and capabilities that might further exacerbate inequalities among genders, communities, journals, institutions, and countries, and (4) the potential risks inherent to opening data and the specific constraints posed by social science data (costs of using a variety of data formats or loss of access to fields, among others). On this basis, we investigated the operational conditions and questions governing when and how to open research data, from the researcher's perspective and from the journal's perspective. Finally, we presented *M@n@gement's* new open data policy.

Our goal with this paper is to open debate with our community around the needs and challenges of open research data in social sciences in general and management and organization studies in particular. We believe that there is an unaddressed paradoxical tension between open science and its commercial uses. We view science as a global public good (Berkowitz & Delacour, 2020), yet the very principles of open science allow reuse by organizations, in particular for performance purposes. We at *M@n@gement* use a Creative Commons CC BY-NC 4.0 license. This means freedom not only to share data and

adapt it under certain conditions (such as giving appropriate credit) but also to use it for strictly noncommercial purposes. We, thus, diverge from the general open science recommendations of no restrictions, which would, thus, include commercial use.

Next, we believe that opening data as much as possible is crucial to efforts to tackle socio-ecological emergencies and widening inequalities. However, we also believe that it takes time and effort to develop research projects, to design and implement DMPs, to publish research in diamond open access journals, and, at the same time, to open data. In that respect, we even more firmly believe in the need for 'sustainable academia' and 'slow science' that we advocated a few years ago (Berkowitz & Delacour, 2020). We would add to that the need to be reflexive and careful about risks of 'subordination of the researched' (Vijay, 2021a, p. 56). This also means exploring and making space for thinking in a way that is different from the dominant order(s), wherever it may come from (Vijay, 2021b). Reflecting on the CARE principles for Indigenous Data Governance not only helps us do precisely that, but also helps us acknowledge that we produce and reproduce power relations and hierarchies of knowledge in the production, uses, and stewardship of research data as well.<sup>9</sup> In that sense, it is our belief that openness is not simply about transparency, reproducibility, or patrimonialization, but is first and foremost about ensuring inclusive and fair (in the sense of just) accessibility to knowledge for all.

## Acknowledgments

The authors thank Anne Bonneville for her guidance and helpful feedback on our data policy, and the participants of the 2022 Congress on 'L'action publique des données & Les données de l'action publique' held in Strasbourg for their comments and input. The authors also thank Lionel Maurel and Astrid Aschehoug from the InSHS for their support.

## References

- Austin, C. C., Bloom, T., Dallmeier-Tiessen, S., Khodiyar, V. et al. (2015). Key components of data publishing: Using current best practices to develop a reference model for data publishing. *International Journal on Digital Libraries*, 18, 77–92. doi: 10.5281/zenodo.34542
- Berkowitz, H. & Delacour, H. (2020). Sustainable academia: Open, engaged, and slow science. *M@n@gement*, 23(1), 1–3. doi: 10.37725/mgmt.v23.4474
- Bonneville, A., Tucci, I., Vion, A. & Giglio, L. (2021). *Données de la recherche: Pratiques et besoins dans un laboratoire pluridisciplinaire SHS* [PhD thesis]. Laboratoire d'Economie et Sociologie du Travail (LEST).

<sup>9</sup> Our inclusive citational practices policy can be found at <https://management-aims.com/index.php/mgmt/code>.

- Borgerud, C., & Borglund, E. (2020). Open research data, an archival challenge? *Archival Science*, 20(3), 279–302. doi: 10.1007/s10502-020-09330-3
- Burgelman, J.-C., Pascu, C., Szkuta, K., Von Schomberg, R. et al. (2019). Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. *Frontiers in Big Data*, 2, 43.
- Carroll, S.R., Herczog, E., Hudson, M., Russell, K. et al. (2021). Operationalizing the CARE and FAIR principles for indigenous data futures. *Scientific Data*, 8(1), 108. doi: 10.1038/s41597-021-00892-0
- Fabre, C. & Gouzi, F. C. (2020). *Proposition de modèle de politique pour les revues et éditeurs quant aux données de la recherche*. Retrieved from <https://hal-univ-tlse2.archives-ouvertes.fr/hal-03026731>
- Feret, R., Gouzi, F. C., Guignon, S., Jouguet, H. et al. (2021). *Guidelines for journals that wish to establish a 'data policy' related to their publications* (10 p.) [Report, Comité pour la science ouverte]. doi: 10.52949/19
- Gauquelin, C., Lutoff, C. & La Branche, S. (2017). Quelle place pour les données issues des sciences humaines et sociales dans le développement de services climatiques régionaux? *VertigO*, 17(3). doi: 10.4000/vertigo.18891
- Hrynaskiewicz, I., Simons, N., Hussain, A., Grant, R. et al. (2020). Developing a research data policy framework for all journals and publishers. *Data Science Journal*, 19(1), 5. doi: 10.5334/dsj-2020-005
- Longley Arthur, P. & Hearn, L. (2021). Toward open research: A narrative review of the challenges and opportunities for open humanities. *Journal of Communication*, 71(5), 827–853. doi: 10.1093/joc/jqab028
- Marlet, O., Baudoin, B., Bernard, L., Bézard, L. et al. (2022). *Le livre blanc du consortium 'Mémoires des archéologues et des sites archéologiques': Guide des bonnes pratiques numériques en archéologie*. Consortium MASA chals-03561376v2, 39.
- Martone M. (Ed.). (2014). *Data Citation Synthesis Group: Joint Declaration of Data Citation Principles*. FORCE11. doi: 10.25490/a97f-egy
- Millerand, F. & Bowker, G. C. (2008). Metadata, trajectoires et 'énaction'. In C. Rosental (Ed.), *La cognition au prisme des sciences sociales* (pp. 277–303). Editions des Archives Contemporaines.
- OECD (2016). Research ethics and new forms of data for social and economic research. OECD Science, Technology and Industry Policy Papers n° 34. doi: 10.1787/5jln7vnpxs32-en
- Osterloh, M. & Frey, B. S. (2015). Ranking games. *Evaluation Review*, 39(1), 102–129. doi: 10.1177/0193841X14524957
- Peng, G., Lacagnina, C., Ivánová, I., Downs, R. R. et al. (2021). *International community guidelines for sharing and reusing quality information of individual earth science datasets*. Open Science Framework Preprints. doi: 10.31219/osf.io/xsu4p
- Roche, D. G., Kruuk, L. E. B., Lanfear, R. & Binning, S. A. (2015). Public data archiving in ecology and evolution: How well are we doing? *PLoS Biology*, 13(11), e1002295. doi: 10.1371/journal.pbio.1002295
- Science Europe. (2021). *Practical guide to the international alignment of research data Management*. Retrieved from [https://www.scienceeurope.org/media/4brkxxe5/se\\_rdm\\_practical\\_guide\\_extended\\_final.pdf](https://www.scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf)
- Stats NZ. (2020). *Nga Tikanga Paihere: A framework guiding ethical and culturally appropriate data use*. Retrieved from [www.data.govt.nz](http://www.data.govt.nz); <https://data.govt.nz/assets/data-ethics/Nga-Tikanga/Nga-Tikanga-Paihere-Guidelines-December-2020.pdf>
- The FAIR Data Principles – FORCE11*. (2014, September 3). Retrieved from <https://force11.org/info/the-fair-data-principles/>
- Vijay, D. (2021a). Vulnerability as praxis in studying social suffering. In E. Bell & S. Sengupta (Eds.), *Empowering methodologies in organisational and social research* (pp. 55–74.). Routledge. doi: 10.4324/9780429352492-4
- Vijay, D. (2021b). Settled knowledge practices, truncated imaginations. *Organization*, 0(0), onlinefirst. doi: 10.1177/13505084211057261
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G. et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. doi: 10.1038/sdata.2016.18
- Willinsky, J. (2005). *The access principle: The case for open access to research and scholarship*. MIT Press.
- Woolston, C. (2021). Impact factor abandoned by Dutch university in hiring and promotion decisions. *Nature*, 595(7867), 462–462. doi: 10.1038/d41586-021-01759-5
- Xu, F., Wu, L. & Evans, J. (2022). Flat teams drive scientific innovation. *PNAS*, 119(23), e2200927119. doi: 10.1073/pnas.2200927119