RESEARCH PAPER

Data sharing policies in scholarly publications: interdisciplinary comparisons

Michal Tal-Socher[1] and Adrian Ziderman[2]*

[1]Consultant, 17 Begin Street, Givat-Shmuel, Israel
[2]Economics Department, Bar-Ilan University, Israel

ABSTRACT
Digital sharing of research data is becoming an important research integrity norm. Data sharing is promoted in different avenues, one being the scholarly publication process: journals serve as gatekeepers, recommending or mandating data sharing as a condition for publication. While there is now a sizeable corpus of research assessing the pervasiveness and efficacy of journal data sharing policies in various disciplines, available research is largely piecemeal and mitigates against meaningful comparisons across disciplines. A major contribution of the present research is that it makes direct across-discipline comparisons employing a common methodology. The paper opens with a discussion of the arguments aired in favour and against data sharing (with an emphasis on ethical issues, which stand behind these policies). The websites of 150 journals, drawn from 15 disciplines, were examined for information on data sharing. The results consolidate the notion of the primacy of biomedical sciences in the implementation of data sharing norms and the lagging implementation in the arts and humanities. More surprisingly, they attest to similar levels of norms adoption in the physical and social sciences. The results point to the overlooked status of the formal sciences, which demonstrate low levels of data sharing implementation. The study also examines the policies of the major journal publishers. The paper concludes with a presentation of the current preferences for different data sharing solutions in different fields, in specialized repositories, general repositories, or publishers' hosting area.

Introduction

This paper addresses current practices of digitally and openly sharing academic research data. More narrowly, it looks into one of the powerful promotion and enforcement tools of data sharing – journal policies encouraging or requiring data sharing. While this research focuses on data sharing policies of journals and publishers, the authors do acknowledge that other tools for encouraging data sharing exist; indeed, in certain disciplines these may have greater heft than journal publication policies. The research was designed to examine interdisciplinary differences in journals' data sharing policies, which are known to exist but not often studied directly. The study is novel in examining data sharing policies in a range of disciplines, and in employing a common research methodology. It also provides additional information on the data sharing policies of the major journal publishers. This empirical research is framed by a more theoretical discussion of ethical issues surrounding data sharing, especially relating to questions of intellectual property rights over research data.

The paper opens with a discussion of the arguments aired in favour and against data sharing (with an emphasis on the ethical issues involved). The next section summarizes available research

***CONTACT**: zidera@biu.ac.il
ACCEPTING EDITOR: Peter Drahos

findings on data sharing by discipline. This is followed by an account of the research method employed. Results are then provided relating, respectively, to journal data sharing policies, to publisher data sharing practice, and data preservation policies. The final section presents some major conclusions emanating from the research.

Data sharing: pros and cons

This section looks at several hindrances to sharing, mainly those that have ethical resonances; we leave aside technical and organizational difficulties that are ultimately more easily resolvable. The calls to share data with the scientific academic community, and with the general public, are made in the name of several values: generally, these are ethical-democratic values and the advancement of science, while some instrumental considerations are also recruited. The opposition to data sharing is more tacit, but several ethical values and a host of instrumental considerations in favour of data withholding can be discerned.[1] We may identify two main groups of hindrances to sharing which have ethical resonance.

Mixed motivations of scientists

Many scientists subscribe to the value of 'communism' (or 'communalism'), the term offered by science sociologist Robert Merton: the ethos of science acknowledges only minimal property rights of scientists over their findings (Merton, 1973/1942, pp.267–78). Data sharing norms, and open science norms generally, clearly derive from this value. Yet, historically informed assessments claim that the question of who is the owner of scientific results and knowledge was never resolutely decided in favour of the public (McSherry, 2001), and also that new fissures formed in this value in the process of commercialization that academia has undergone, especially since the 1980s (Macfarlane and Cheng, 2008; Radder, 2010). Yet again, a counter movement is visible which strives to preserve the classical scientific ethos by creating an array of new means of enforcements (Montgomery and Oliver, 2009).

Aside from a general commitment to communalism, in many fields researchers benefit greatly from access to the research data of others (National Research Council, 1997; Brown, 2003). Yet, at the level of the individual researcher, sharing norms may run counter to self-interest, at least in the short run. This is because it is time-consuming and not well rewarded, as publishing data is not given credit on a par with published papers. Recognition is very important as it is also ethically acceptable that scientists see themselves to be true owners of the data they have collected with effort, time, and talent. Thus, in many cases, it may be argued that data should not be treated simply as a public raw resource (Mauthner and Parry, 2013).

Important steps toward changing the equation have been made in recent years, and now citation standards are being devised; many repositories coin persistent digital identifiers (long-lasting links) for datasets hosted by them, and many publishers enable linking papers to these datasets (Callaghan, 2014). Also, metrics for data, similar to the impact factor, are being developed and implemented.[2] But further institutional changes will have to take place in order for funding bodies and university promotion committees to acknowledge shared data as academic achievements.

Other concerns are even more difficult to alleviate: researchers may sacrifice their competitive edge by sharing data before exhausting research possibilities, and may abhor exposing themselves to further criticism and misinterpretation. Probably, these difficulties can be handled

[1] It should be borne in mind that data sharing in the more taxing sense it has today is a first-world preoccupation. Researchers from lower income countries are far from able to participate as data donors (e.g. Rappert and Bezuidenhout, 2016).

[2] E.g. the data citation index of Clarivate Analytics (formerly, Thomson Reuters) available at http://wokinfo.com/products_tools/multidisciplinary/dci/about/ (accessed May 2020).

only when everyone is expected to share, and some sanctions may be unavoidable on the way to consolidating these habits, as is often the case when new norms are being established. Lastly, researchers may feel more committed to familiar others, such as colleagues (and their self-interest) and respondents (privacy issues), than to other somewhat abstract others. This may also be considered ethically acceptable (Mauthner and Parry, 2013). Some repositories allow greater control of depositors over the level of publicity of all or some of the data. Procedures of de-identifying data are also being used in repositories hosting data from human subjects, although these are not without controversy (Kaye, 2012).

Commercial pressures

As universities and research institutes become increasingly exposed to economic pressures, and in certain regions are expected to be more self-supporting, they understandably strive to secure intellectual property rights on research results (Reichman and Uhlir, 2003, pp.341–2). This is a major issue beside the more basic ethical question of who owns research results, which historically was sometimes the institutions creating knowledge (McSherry, 2001). Private companies are also interested in promoting a protective intellectual property rights regime for publicly funded research as this enables them to acquire exclusive licences to complete R&D toward marketability. Fittingly, legal changes have been introduced to facilitate protecting intellectual rights of publicly funded research results, now sometimes extendable to data (Reichman and Uhlir, 2003, pp.315–462). That said, for some policy makers, one of the justifications for data sharing is driven by economic considerations resonating with free market values. Generally, they argue that in the bigger picture nations can benefit from open research data whose commercialization occurs more downstream (e.g. Arzberger *et al.*, 2004). These two sets of hindrances, which cannot be easily removed and which have certain ethical and instrumental backing, are reflected in data sharing norms being far from universally accepted.

Related work

Academic publishing roles in data sharing

The publishing process has become an important avenue for promoting and enforcing research integrity norms because it is in a position to offer clear incentives for compliance, being the pivot of the academic reward system for researchers (Sturges *et al.*, 2015). These norms are promoted and sometimes enforced by individual publishers, journal editors, and societies which publish journals, or by voluntary consortia of academic journal editors which strive to shape norms for whole disciplines by promoting guidelines and standards (Montgomery and Oliver, 2009). This is also true for data sharing norms. First, consortia of editors occasionally play a role in setting data sharing norms. The statement by political science journal editors promoting the adoption of data sharing practices is a case in point (Bonneau *et al.*, 2015). Second, many publishers and journal editors (and also societies) have all kinds of policies and statements regarding the handling of data that underlie the research reported in submitted papers.

It may be noted at this stage that whereas the publishers' role in promoting and enforcing open access to research papers is clear, their expected involvement in data sharing is much less so. Beyond promoting sharing by policies, journals and publishers are sometimes involved in data sharing as hosts: as we shall see, it is still a common practice for journals to host data related to papers published in them on the publisher's website. However, this solution has many shortcomings and some journals already find it hard to handle the large amounts of data thus generated (Dallmeier-Tiessen *et al.*, 2012, pp.32–7). In addition, it is not entirely clear and agreed what are the responsibilities of the journals in quality checking the datasets deposited (Beagrie, 2010).

In the case of publishers' business models supporting public sharing, the solution is open access to research papers. Authors (or their sponsors) pay fees to cover publishing costs for open access papers instead of the traditional model of library or reader payments. This in turn has created a new barrier for researchers from lower income countries to publishing their research results (e.g. Zachariah *et al.*, 2014).

In the case of data sharing, the problem is even greater, as the responsibility is diffused among more players, both publishers and repositories. If data are hosted by publishers, then publishing or accessing them is restricted, as is the paper related to the data. If data are hosted by repositories, then deposition and accessibility depend on the policy of the repositories (Campbell, 2015; Assante *et al*., 2016). Another cost relates to quality checking the data: even if data do not go through a proper peer review, which occurs today only very rarely, at least checking that files are accessible, full, and understandable is called for. For journals, in order to truly promote data sharing, this checking should be undertaken and the costs will need to be covered in some way. At present, when data sharing norms are far from being universal, researchers have little motivation to pay additional fees either to journals or to repositories for depositing or quality checking their data. It remains to be seen whether a viable business model will be found for publishers that truly supports open access to data (Vines, 2017). The business model of a group of repositories can be viewed at Dataverse (2017).

Previous research on journal policies

Many available studies have assessed the needs for, and objections related to, data sharing. Some of these studies also tapped publishers' and editors' perceptions and self-reported practices (Kuipers and van der Hoeven, 2009; Beagrie, 2010; NISO, 2013; Sturges *et al.*, 2015). Other studies have tried to assess actual levels of data sharing, relying on researchers' self-reporting (Swan and Brown, 2008; Kuipers and van der Hoeven, 2009; Tenopir *et al.*, 2011; Tenopir *et al.*, 2015).

An alternative approach has been adopted in other cases, studying data sharing policies declared by journals, usually in the 'instructions for authors' section. At a minimum, these studies give an indication of the stated importance and prevalence of data sharing norms. Perhaps the earliest study of this sort surveyed 850 journals (McCain, 1995); at that time, the prevalence of policies was very low (about 15%). A 2015 study examining about 370 journals found that about half had data sharing policies (Sturges *et al*., 2015). The earlier study also indicated that biomedical journals showed the highest adoption rates of data policies.

A much discussed question relates to the usefulness of adopting stronger policies by which papers would be accepted for publication only if certain types of data are publicly shared prior to publication. Making data sharing mandatory in this way is unlikely to be sufficient in itself to overcome researchers' motivation barriers, lack of appropriate digital infrastructure and other obstacles (Nelson, 2009; Crotty, 2016). However, stronger policies do have a positive effect on sharing rates: mandating sharing for some types of data is a necessary though not sufficient component for a successful sharing culture. Thus, a study that examined 70 journal policies in an area of biomolecular studies found that policy strength was positively associated with actual data sharing (Piwowar and Chapman, 2008; see also Vines *et al.*, 2013). A study examining changes in data sharing policies of molecular biology journals over 20 years showed policies becoming more demanding and mandatory, with an associated rise of actual data sharing (Brown, 2003). Another study of data sharing policies of economics journals argued strongly for mandating sharing even though there are no widely accessible specialized repositories in the field and no broadly accepted standards for sharing (Vlaeminck, 2013). However, experience teaches that rules too distant from common norms usually fail (e.g. Bicchieri and Mercier, 2014). Thus, introducing mandatory policies in research domains without community support for sharing norms is unlikely to succeed.

Interdisciplinary differences in data sharing

There are a few comprehensive studies that describe and compare the different motivators, characteristics and needs involved in data usage that is characteristic of various disciplines and research communities (Harley *et al.*, 2010; PARSE Insight, 2010). These case studies of multiple disciplines also discuss data sharing and have proved to be an illuminating source for understanding interdisciplinary differences in sharing. Together with more limited studies on data sharing, they show the following outlines:

- ***Biomedical Sciences*** differ among themselves in the volumes of data produced, molecular biology being a very data intensive field (Stephens *et al.*, 2015). Some biomedical fields have cultivated the most advanced data sharing culture (Hu and Kaabouch, 2014, p.181). Appropriately, most research on data sharing focuses on biomedical sciences (Fecher *et al.*, 2015).
- ***Physical Sciences*** manage massive volumes of data; some, such as astronomy, were pushed relatively early to adopt practices of data sharing (Hu and Kaabouch, 2014, p.182). Many projects fall into the category of Big Science (very large teams, highly instrumentalized, very expensive), which is more prone to adopt data sharing (Reichman and Uhlir, 2003, p.322). Yet, over all, sharing rates are lower than those in the biomedical sciences (e.g., McCain, 1995).
- ***Social Sciences*** – several social science disciplines were among the earliest to organize efforts to share data, especially political science, sociology, and economics (Harley *et al.* 2010, pp.677–83; Pienta *et al*., 2010). There is great variance among the different social science disciplines, the more empirical ones adopted sharing much more than the hermeneutical ones (PARSE Insight, 2010, p.58; Pienta *et al*., 2010). Growing data intensity is very much felt in some types of research (Elias and Entwisle, 2013). Overall, however, social science data are not being shared (Pienta *et al*., 2010).
- ***Arts and Humanities*** – the data sharing infrastructures for the humanities are much less developed than those created for some of the social sciences (IFDO, 2014). Again, more empirical disciplines, such as linguistics and archaeology, are performing somewhat better (Harley *et al.*, 2010, pp.29–139; PARSE Insight, 2010, p.58) and progress has been made recently even in other fields, such as ethnomusicology (Harley *et al.*, 2010, pp.570–5), though this is probably still rather marginal.
- ***Formal Sciences*** such as mathematics and computer science are rarely discussed in relation to data sharing, separately and even less so as a group. Sometimes they are considered as part of the physical sciences, though their non-empirical nature should keep them apart. Indications are that they do not adopt the sharing culture of some physical sciences (McCain, 1995; for a glimpse on the different data usage patterns in mathematics see Womack, 2015).

There are, however, very few studies comparing journal data sharing. The early study by McCain (1995) found that about two-thirds of the journals with sharing policies belonged to the life sciences, while the rest came from physics and chemistry. Journals in medicine, engineering, mathematics, and computer sciences essentially had no such policies. No other disciplines were examined in this rather rare interdisciplinary study.

Other findings are available from studies with a narrower disciplinary breadth. As noted, in the biomedical sciences mandatory data sharing policies started to emerge in the late 1980s, and the trend has been strengthening ever since.[3] Among the social sciences, only in economics is there a

[3] See e.g. the history of announcements by ICMJE regarding clinical trial registration, available at www.icmje.org/news-and-editorials/ (accessed May 2020).

clear trend of journals adopting mandatory data sharing policies, though inaugural efforts are being made in political science and sociology (Pienta *et al*., 2010; Zenk-Möltgen and Lepthien, 2014). There do not appear to be any similar studies in the arts and humanities, although there are archaeology journals with data sharing policies. Nor are we aware of any study examining formal sciences journal data sharing policies (though mathematics journals with such policies are included in our journal sample).

The research reported in this paper attempts to add to the relative paucity of research into interdisciplinary differences in journal policies. The study does so though a cross-disciplinary comparison using a common methodological framework. It also includes disciplines that are rarely (if ever) studied in this context, notably in the arts and humanities.

Materials and methods

Definitions of data sharing

The websites of 150 journals and of major journal publishers were examined for information on data sharing policies. Two definitions of data sharing were adopted for this purpose:

- ***Enabling data sharing*** refers to policies where data sharing is possible, but not mandatory. It is defined as the sharing of research data underlying published academic research papers, on a digital platform accessible to the general public or to an entire research community (although restrictions may be in place) and linkable to the paper in question. Neither the sharing of data with individuals upon private request nor sharing related specifically to the review process is included in this definition.
- ***Strong data sharing*** is in place where at least some types of data must be deposited for open sharing as a condition of publication.[4] A dilemma facing the authors was posed by clinical trials registration (CTR) in that some publishers and journals publishing health-related papers mention it as the only data type whose sharing is mandatory. CTR is somewhat different from other types of data treated in the publication process as it does not refer to data supplementing the information presented in the paper. Proper registration is made long before publication.[5] It was decided to report separately on journals with strong policies that excluded and included those which only mandate CTR.

It should be noted that 'open sharing' has a wide spectrum of meanings for different hosts and depositors of data. In principle, this usually means that data should be available to all. However, sometimes embargos are imposed, access restricted to certain user groups, and parts of the dataset defined as confidential (see especially Eschenfelder and Johnson, 2014). Even when unrestricted, access is not always anonymous and sometimes requires registration. The terms of use of deposited data are not always stated clearly and the data may be covered by laws prohibiting certain types of reuse. Sometimes licences for reuse are stated clearly, but here too there is a variety of options, even though many licences are quite permissive (similar to CC0 or CC-BY licences). Curation activities of hosts also vary widely, both relating to technical and general services as well as to more content-specific services, which cannot be offered by more heterogeneous hosts (Assante *et al*., 2016; Austin *et al*., 2015; Ball 2014; Campbell, 2015; Dataverse, 2017).

---

[4] Our definition is similar to the definition of strong policy in Piwowar and Chapman (2008), yet broader in order to apply to more data types and preservation solutions.

[5] ICMJE's widely used definition of clinical trials is to be found at www.icmje.org/about-icmje/faqs/clinical-trials-registration/ (accessed February 2015).

Building the sample

Academic journals differ in quality and prestige. In order to set up a representative sample of journals for the study, we first chose to use the SJR (SCImago journal rank) prestige metric as an indicator of journal influence. The SJR ranks journals included in the Scopus bibliographic database. We then chose fifteen disciplines from five main categories of academic disciplines and used the Scopus classification titles for them. The number of indexed journals in the different categories varies widely, so to reduce the size of very long lists to make them comparable with the others, representative subcategories were sometimes chosen. The categories used are shown in Table 1; the column to the right specifies reasons for choosing these particular disciplines.

In all, 150 journals were sampled, ten journals for each discipline. Thus, unlike the usual research strategy in this field, we opted for a broader sample distribution at the expense of depth. Each disciplinary list is ordered according to SJR scores. For each discipline, we chose journals from the entire range: two of the top five journals; two around the end of the 25th percentile; two around the median; two around the end of the 75th percentile; and two of the lowest ten. Former studies have shown that journals with higher influence metrics more often have data sharing

Table 1. Disciplines represented in the study

Discipline group*		**Discipline**	**Reasons for choosing the discipline**
Biomedical Sciences	Life Sciences	Genetics	Much studied in data sharing context – the paradigm for data sharing
		Neuroscience	Newer life science
	Medicine	Oncology	Involves humans (privacy concerns); heavily invested, competitive
		Pharmacology (medical)	Does not necessarily involve humans
Physical Sciences		Chemistry (miscellaneous)	More theoretical
		Geology	Data rich as a highly automated observational science
		Ecology	Public interest and participation
Social Sciences		Economics	Heavily quantitative, data rich
		Social Psychology	Theoretical aspects, empirical and experimental aspects, clinical aspects
		Political Science and International Relations	More hermeneutic than the above
Arts and Humanities		Archaeology (Arts and Humanities)	Known to have specialized repositories; partly empirical
		Music	May be data rich; partly empirical and partly hermeneutic
		History	More hermeneutic than the above
Formal Sciences		Computer Science (miscellaneous)	More applied; itself involved in developing tools for sharing; very little is known
		Statistics and Probability	More theoretical; very little is known

*Titles of the main category groups are not taken from Scopus policies and tend to have stronger policies (probably as they are better positioned to set higher standards). So, in order to have a balanced picture, it is important to keep the sample representative in this respect (Piwowar and Chapman, 2008; Vlaeminck, 2013; Sturges *et al.*, 2015).

The SJR indicator was preferred over the more pervasive Clarivate (formerly Thomson Reuters) impact factor because the latter is based on the Web of Science database, which is generally much less comprehensive – 13,500 journals in WoS versus about 20,500 journals in Scopus (Mongeon and Paul-Hus, 2016). SJR is also considered a better indicator of a journal's influence as it gives greater weight to citations coming from more prestigious journals. Additionally, the impact factor is arguably more amenable to manipulations by journal editors (Ramin and Sarraf Shirazi, 2012). We used the SJR data for 2014 (counting 2014 citations of citable papers appearing in the three preceding years) as retrieved in October–December 2015.

Many journals appear in several disciplinary lists, but it was verified that the discipline we wanted the journal to represent had more than tangential relevance to the journal's scope. Journals publishing only reviews were generally avoided (as research data are much more likely to arise from original research). We avoided including two journals from the same publisher for a given ranking area (i.e., top five, median, etc.). It was verified that the journal was covered continuously in the years relevant for the 2014 SJR score, and that it had no fewer than 30 citable documents in these years. It was sometimes difficult to find end-of-the-tail journals that meet all these criteria. In these cases, journals higher on the list than the last ten were chosen (usually the differences between SJR scores in this region are very small) or, if no better candidates were found nearby, one of the criteria was compromised. All these procedures, while admittedly unblinding the selection process, were applied in order to make the small group of journals representing each category more informative. Non-English websites were not shunned, Google Translate being employed for websites in languages other than English, German, or Italian. Since the size of the sample for each discipline is small, our results may not be fully representative. Yet the manageable size did enable us to keep the detailed policies and their qualitative dimensions in sight. The results for the disciplinary groups are informative and lead to several insights.

Data collection

For each journal, the journal's website was searched to see whether there was a text instructing authors on the sharing of data supporting the research presented in the submitted paper. Such texts are usually found in the 'Instructions for Authors' section, but are sometimes incorporated into an ethical policy section or given a space of their own. Where a data sharing policy was found, it was recorded in terms of the nature of the guidelines indicated. These included whether sharing is only encouraged or required, and whether it is intended for the review process only, for public/community open sharing or for sharing upon private requests. The types of shareable data were recorded, where data should be shared, when data should be shared, and whether there were procedures for exemption from sharing. For each journal, we also recorded the publisher (or specific publisher's imprint) and publisher type (for profit or not) and looked to see whether the publisher had a data sharing policy of its own. These are found either on the journal's webpage, where it is made clear that the publisher is the source of the policy, or on the publisher's general website.

Publishers' policies were analysed similarly to journal's policies. Additionally, for each journal we looked for a related academic or professional society involved in its publication. If there was one, the society's website was searched to see whether it declared a data sharing policy of its own. If present, the policy was analysed using the same categories. For each journal, we indicated whether open data sharing is at all possible and, if so, whether it is data sharing in the strong sense, regardless of the source of the policy (journal, publisher/imprint, related society).

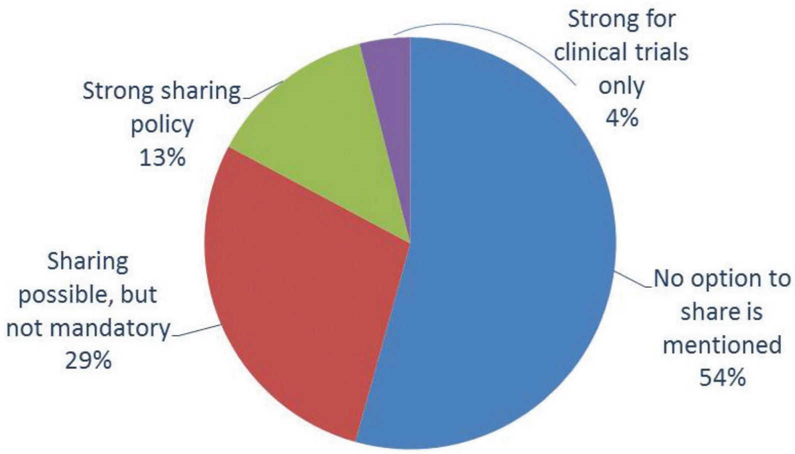

Figure 1. Journal data sharing policies

Results

Journal data sharing policies

Possible and strong sharing

Sixty-nine out of the 150 journals included in our study (46%) enable data sharing one way or another. Of the 69 journals that allow for data sharing, only 20 journals have a strong policy according to which, at least *de jure*, papers will not be published unless certain types of data are deposited. This means that only 13% of the journals in the sample have a policy with a fair chance of being followed (Figure 1).

The figures in the current study are slightly higher if we consider journals that require CTR to be strong policy journals – 26 journals, 17.3% of our sample, have strong policies according to this wider definition. Yet, almost half the journals in the sample offer authors an opportunity to share data, which means that at least some sharing infrastructure is already in place for a considerable share of academic publications.

The study confirmed that, overall, journals with higher citation metrics more frequently adopt data sharing policies. It was found that about 74% of the journals enabling sharing have SJR scores from median to the top of the distribution, and 80% of the journals with strong policies are located in the upper halves of their discipline lists (compared with the neutral expectation of 50% in both cases). This is generally true also, with some variations, at the level of the disciplinary group. The formal sciences, alone, do not fit this pattern: only 57% of journals enabling sharing are in the upper half; and the single journal with strong policy is in the lower half (see Table 2 for details).

Interdisciplinary differences

This global picture of the pervasiveness of data sharing policies changes considerably as we look at the level of individual disciplines. To make these tendencies more robust, we also look at differences among discipline groups (biomedical sciences, etc.). These figures are detailed in Table 3 (where data sharing is possible) and Table 4 (for strong data sharing policies). The results for the discipline groups point to several conclusions.

- ***Biomedical Sciences*** lead the trend of data sharing, in line with what is known from earlier research. In all the four disciplines representing the field, over 50% of the journals enable sharing, which means that solutions for sharing in this area are quite widespread. Yet *requiring* data sharing for at least some types of data is less common (only 30%, disregarding policies merely demanding CTR). The ratio strong/possible is especially low for the two

Table 2. Pervasiveness of the possibility to share data: interdisciplinary comparison

Discipline Group	**Discipline**	**Individual discipline**		**Discipline group**	
		Journals enabling sharing No.	**Journals enabling sharing* %**	**Journals enabling sharing No.**	**Journals enabling sharing* %**
Biomedical Sciences	Genetics	8	80%	27	67%
	Neuroscience	6	60%		
	Oncology	6	60%		
	Pharmacology (medical)	7	70%		
Physical Sciences	Chemistry	4	40%	13	43%
	Geology	5	50%		
	Ecology	4	40%		
Social Sciences	Economics	5	50%	14	47%
	Social Psychology	6	60%		
	Political Science and Int. Relations	3	30%		
Arts and Humanities	Archaeology	4	40%	8	27%
	Music	2	20%		
	History	2	20%		
Formal Sciences	Computer Science	3	30%	7	35%
	Statistics and Probability	4	40%		
Total		69		69	

*Out of the10 journals that were sampled for each discipline

Table 3. Data sharing differences among discipline groups

Discipline group		**Data sharing is possible**			**Strong data sharing**		
	No.	**Total**	**In upper half**		**Total**	**In upper half**	
		No.	**No.**	**%**	**No.**	**No.**	**%**
Arts and Humanities	30	8	8	100	1	1	100
Biomed	40	27	18	67	12	10	83.
Formal Sciences	20	7	4	57	1	0	0
Physical Sciences	30	13	11	85	3	3	100
Social Sciences	30	14	10	71	3	2	67
Total	150	69	51	74	20	16	80

medical disciplines (disregarding CTR-only policies: 2/6 for oncology, 1/7 for medical pharmacology). Previous studies have similarly shown that research more oriented to medical purposes is less shared (McCain, 1995; Piwowar and Chapman, 2008; Milia *et al.*, 2012).

- ***Physical Sciences** and **Social Sciences*** demonstrate similar levels of adoption of data sharing policies, both at the level of enabling sharing (43% and 47% respectively, with somewhat greater inner variance for the social sciences) and at the level of strong policies (10% both).

Table 4. Pervasiveness of strong data sharing policies – interdisciplinary comparison

Discipline Group	Discipline	Individual disciplines		Discipline group	
		Journals with strong policy No.	Journals with strong policy+ %	Journals with strong policy No.	Journals with strong policy+ %
Biomedical Sciences	Genetics	5 (5)*	50% (50%)	12 (18)	30% (45%)
	Neuroscience	4 (4)*	40% (40%)		
	Oncology	2 (5)*	20% (50%)		
	Pharmacology (medical)	1 (4)*	10% (40%)		
Physical Sciences	Chemistry	2	20%	3	10%
	Geology	1	10%		
	Ecology	0	0%		
Social Sciences	Economics	2	20%	3	10%
	Social Psychology	0	0%		
	Political Science and Int. Relations	1	10%		
Arts and Humanities	Archaeology	1	10%	1	3%
	Music	0	0%		
	History	0	0%		
Formal Sciences	Computer Science	1	10%	1	5%
	Statistics and Probability	0	0%		
Total		20 (26)*		20 (26)	

+ Out of the 10 journals that were sampled for each discipline

* In parentheses – including journals with strong policies only for clinical trials

This may appear surprising as there are more specialized data repositories for different types of physical data. Results may have looked different for other specific disciplines. Astronomy, for example, is considered paradigmatic for data sharing as data from national facilities are often shared (Harley *et al*., 2010, pp.170–3; Hu and Kaabouch, 2014, p.182). Earth observation methods are also known to have specialized repositories, but the two disciplines in our study utilizing these methods, geology and ecology, do not embrace data sharing comprehensively enough to leave a mark on our sample.

- ***Arts and Humanities*** lag behind the other academic meta-fields in terms of the adoption of data sharing norms. Only about 27% of the journals in this category enable sharing at all; just a single journal requires the deposition of data as a condition for publication.
- ***Formal Sciences***, about which little could be gleaned from the available literature on data sharing patterns, are characterized by a very low level of data sharing norms adoption (only 35% of the journals enable sharing, only one journal in the sample had a strong policy of data sharing).

Publisher data sharing practice

The scientific publishing industry is currently undergoing a process of centralization. In parallel, for-profit publishers are becoming more dominant, while the role scholarly societies play in publishing is becoming more secondary (Larivière *et al*., 2015). These trends are also reflected in our sample. First, this may explain why the data sharing policies we encountered in our sample rarely

Table 5. Publishers with most journals in the current study and the Scopus database

	Current study			**Scopus**	
	Rank	**Journals**		**Rank**	**% of sources in Scopus (2015)****
		No.	**%**		
Taylor & Francis	1	15	10%	4	4.3%
Elsevier	2	13	9%	1	10.4%
Springer	3	11	7%	2	7%
Wiley-Blackwell	4	9	6%	3	4.4%
Nature Publishing Group*	5=	4	3%	X*	X*
Nova Science Publishers	5=	4	3%	72	0.02%
Sage	6	3	2%	5	1.7%

*Calculated as part of Springer in Scopus database due to a recent merger, so independent ranking data are not available

**Out of 11,441 sources

Table 6. Share of Scopus top five publishers in data sharing policies, within sample

Publisher	**Journals enabling sharing No.**	**Journals with strong sharing policy No**
Elsevier	13	6
Springer	9	2
Wiley-Blackwell	6	3
Taylor & Francis	8	0
Sage	2	1
Total	38	12
%	56% *	46% **

*Of all journals enabling sharing in sample

**Of all journals with strong policy in sample

originated in societies (two cases or 10% of the strong policies) and reflected much more often requirements of single journals (50% of the strong policies) or publishers/imprints owned by publishers (40% of the strong policies), which seem to have greater power to further these demands.

Second, Table 5 shows the publishers sampled most frequently, compared with the top publishers represented in the Scopus database from which all the sample's journals are taken.[6] The levels of concentration evident from this table justify focusing on the top publishers. As can be seen, our sample is quite representative of the general Scopus database for the top four publishers, which together publish 26% of the entire publications in a database that enumerates more than 5,000 publishers (afterwards biases are introduced, among other things, because of the different proportions of the disciplines in the sample and the entire database). All the publishers in the table are commercial publishers.

All major Scopus publishers noted in Table 6 – Taylor & Francis, Elsevier, Springer Nature, Wiley-Blackwell, and Sage – have mechanisms in place to enable data sharing. Some of these publishers host supplementary materials on their websites, some enable linking to approved outside repositories, and some do both. Not all of them administer peer review for data hosted in house.

[6] Data on the publishers on the Scopus database were derived from the journal title list file downloadable from www.elsevier.com/solutions/scopus/content.

Table 6 shows that these top five publishers account for 56% of the journals enabling sharing in our sample, although they publish only 34% of the sample journals. They also publish 46% of the strong policy journals, again somewhat more than neutrally expected from their share in the sample.

None of these big publishers, which run very heterogeneous portfolios, endorses a strong policy obliging all its journals to share data (though Wiley-Blackwell does require editors of medical journals to insist on prospective CTR). Some imprints belonging to these publishers do have stricter requirements (e.g. *Cell*, now owned by Elsevier). Similarly, there are smaller publishers with more homogeneous STM portfolios that also require some types of data to be deposited as a condition of publication (e.g. Nature Publishing Group, now acquired by Springer).[7]

Data preservation solutions

Hierarchy of preservation solutions

A closer look into the data sharing policies encountered in biomedical science can indicate what, currently, are the pervasive solutions for sharing. All the twelve journals with strong policies (disregarding those with strong policies for CTR only) require the deposition of some types of data in public, community-endorsed, curated, specialized data repositories for specific data types (e.g. GenBank for DNA sequences, wwPDB for structures of biological macromolecules). Specialized repositories are clearly the first preference. When these do not exist (or still do not enjoy broad endorsement), journals and publishers often encourage, or require, authors to submit other types of data to the supplementary material area of the journal and/or to general scientific data repositories that can serve many types of data (e.g. FigShare, Dryad).

It is not always clear whether supplementary material submitted to the journal's website rather than specialized repositories is peer reviewed and checked for user- friendliness and usefulness (whether it is presented clearly with all the metadata required; whether it is searchable and can be analysed easily, whether it can be clearly cited), and whether long-term preservation is ensured (Kuipers and van der Hoeven, 2009, pp.50–3; Beagrie, 2010). The general-purpose repositories, such as Dryad and FigShare, offer a much less specialized quality control, and the services given by different repositories vary widely (e.g. Austin *et al*., 2015). Yet, some of them offer much better long-run preservation solutions (such as being more citable, searchable, and other archival services) than those offered by journals.

As to the eight cases of journals with strong policies outside the biomedical sciences, none of these requires the deposition of certain data types in external specialized repositories. At most, these are mentioned only as optional (e.g. the *Journal of Anthropological Archaeology* mentions the possibility of depositing data in specialized archaeological repositories). General-purpose repositories are only mentioned once (Dataverse is the mandatory host for at least some types of papers published in Economics). The default hosting solutions for these journals are journal/publisher websites, with all their shortcomings. Among the non-biomed journals enabling but not requiring sharing, this is also the preferred solution. Specialized and general external public repositories are mentioned only twice out of 34 cases.

Preservation patterns across disciplines

Of the biomedical journals examined, only *Genome Research* requires full disclosure of data. Much more commonly, biomedical journals insist on certain types of data only. The frequency of the requirements for each data type can instruct us on its current level of institutionalization. In biomedical sciences, as represented in our research, the most institutionalized data types are clearly

---

[7] Springer Nature has recently made an effort to standardize the data sharing policies of its journals, offering its editors a choice of four possible policies. According to the latest update, about 70% of the journals that entered the process adopted one of the two weaker policies (Hrynaszkiewicz *et al.*, 2017).

related to genetics, a pioneering field in data sharing with an embattled history in this area (Jasny, 2013). Even here, different types of data enjoy different levels of sharing. The prominence of genetics-related repositories in our research undoubtedly has to do with the choice to include genetics as one of the four biomedical disciplines, but this is far from the whole explanation. We have included neuroscience as well, a younger discipline with a shorter history of data sharing. Though several specialized repositories exist for certain types of neuroscientific data (such as OpenfMRI), not a single journal in our sample mandates their use.

The picture is different for the four social sciences and arts and humanities journals with strong policies that were sampled. All of them require a wide range of data types. These data types are not only quantitative and structural, but include, for example, instructions given to subjects of experiments and the programs used for running the experiment. It is striking that no repository specializing in the curation of qualitative data was mentioned in the policies sampled, attesting to the slower evolution of these solutions.

Discussion

The results of the study lead to several conclusions regarding the role of publishers in data sharing. It was found that hosting data on a publisher's website is still a widespread solution for sharing. This may have a positive effect: as publishers have already developed some infrastructure for sharing anyway, journals of disciplines which currently have low levels of data sharing may enjoy the spillover of existent infrastructures. It remains to be seen if this has value in the absence of a much more significant effort to promote data sharing in these disciplines. Additionally, hosting by publishers is problematic in the sense that data stored by publishers may be less open to all than data shared in repositories (because of fee barriers for deposition or access). Also, it was found that the top publishers may be considered promoters of data sharing as they have all been pushed to adopt some sort of solution for sharing. Yet their heterogeneous portfolios (in terms of content type and quality) seem to stand in the way of using their considerable power to further strong policies.

We turn now to the data sharing policies across various academic disciplinary groups. A major, confirmatory, finding is that the biomedical sciences lead in data sharing, as is well established in the literature. Yet, medical journals are less likely to have strong data sharing policies than biological journals. This is an indication that the mere existence of specialized and widely endorsed repositories is not, in itself, sufficient to ensure sharing. Many other factors are involved. One such factor is the need to protect the privacy of subjects whose personal data might be traced despite anonymization (this factor also plays a role in social sciences dealing with human subjects (Eschenfelder and Johnson, 2011; Weller and Monroe-Gulick, 2014). Another factor, especially relevant for medical research, is that this field is supported relatively heavily by private investment. The commercial interests of universities and companies wishing to profit from research may deter data sharing (Reichman and Uhlir, 2003, pp.315–462; Eisenberg, 2006).[8]

Similar levels of policy prevalence were found for the physical and social sciences, although many specialized repositories exist for Big Science projects in which some data types must be shared. Perhaps physical data are much less shared outside these projects. Though it seems that the sampled physical sciences could benefit greatly from data sharing (at least in chemistry and geology), this is not the case, and the obstacles may be more normative than technical.

---

[8] Eisenberg's paper relates to biomedical research in general, but all the examples are taken from health-related cases, presumably because the main battlefield is medical research. The author also points to the very complex economic reality created in the last couple of decades surrounding scientific research, where many types of claims for intellectual property rights may be attached to datasets, sometimes with paradoxical results (e.g. cases in which publicly funded scientists avoid sharing data in an attempt to prevent companies from using them and then patenting research results).

At least in some of the social sciences journals do have a regulatory role for data sharing. It is unclear whether social science journals will follow the policy trajectory charted by biomedical journals, in which policies grow stronger for specific data types hand in hand with the consolidation of specialized repositories for these types. The social science strong policy journals sampled here demonstrate a different pattern – they do not require deposition in outside specialized repositories and do require a wider variety of data types. If this observation is valid, and with an eye to the shortcomings of non-specialized repositories, will the journals fulfil a more central role in quality control for these domains, or will other mechanisms be found to ensure it? The existing trend in economics, also echoed in our sample, to make data sharing mandatory in the absence of specialized repositories, should be studied carefully for the viability of this possible alternative evolution of solutions and policies.

The arts and humanities demonstrate the lowest prevalence of data sharing policies. This is probably related to the slower adaptation to digital technologies, with the divergent epistemic culture in which, usually, each researcher is something of an island, and with the exceptional hardships involved in sharing qualitative data that can be multilingual, historically specific, and ambiguous (see Unsworth, 2006). Considerable expertise is required to interpret the data gathered by others (PARSE Insight, 2010, pp.11–41). It may also be that in this domain journals have weaker regulatory power, books being more important than papers for academic promotion (Harley *et al*., 2010).

The formal sciences also demonstrate a low prevalence of data sharing policies. The existing literature did not suggest explanatory factors, and no good description of data usage habits was found in it. These sciences are more often discussed as the forgers of tools for data intensive science (e.g. Baker *et al*., 2010), but their own usage patterns remain in the dark. One factor influencing policy prevalence could be a weaker position of journals to promote norms in the field (Harley *et al*., 2010). These results are another indication that quantitative methods and digitally born/easily digitizable data, in themselves, are no guarantee of data sharing.

Some limitations of the research should be noted. The humble size of the sample may limit the representativeness of the study, though it does provide a rare cross-section view of the field. While this type of study tells us little about actual rates of data deposition, it is informative about expected and evolving norms. Studies that strive to evaluate actual rates of sharing attest to low rates even for papers published in journals with strong policies. In other words, policies are not always enforced (e.g. Piwowar and Chapman, 2008; see also Savage and Vickers, 2009; Alsheikh-Ali *et al.*, 2011; Vlaeminck, 2013). Lastly, data deposition may say little about the usefulness of the data deposited or whether the information is full and clear enough to be used by others. Studies using shared data to try to replicate original results could give an indication of the usefulness of sharing (McCullough *et al.*, 2008; Ioannidis *et al.*, 2009; Atici *et al.*, 2013). Such efforts can obviously be made only by insiders dedicating research to specific disciplines.

Lastly, we consider what guidance the research can offer policy making. While the study cannot resolve the uncertainty of policy makers about whether strong policies are always helpful, it can show them where to look for empirical evidence. A decision in this regard could take into consideration the two approaches taken in our research. Strong policies for particular data types in the biomedical sciences seem to be helpful. But is the alternative model for the social sciences, of strong policies for diverse data types, in the absence of specialized repositories, a viable alternative or a dead end? It is still unclear whether certain disciplines that can benefit, or already benefit, from specialized repositories (such as linguistics and archaeology) could embrace one of these alternatives.

Acknowledgements

The research was sponsored by the Committee on Publishing Ethics (COPE) and was conducted within the framework of COPE's research subcommittee. The conclusions and views expressed are those of the authors and are not necessarily those of COPE. The very helpful comments of the responsible editor and of two anonymous referees are gratefully acknowledged. Some of the findings were presented at the Eighth International Congress on Peer Review and Scientific Publication, September 2017, in Chicago, Illinois.

References

Alsheikh-Ali, A., Qureshi, W., Al-Mallah, M. and Ioannidis, J. (2011) 'Public availability of published research data in high-impact journals', *PloS One*, 6, 9, e24357.

Arzberger, P., Schroeder, P, Beaulieu, A. *et al.* (2004) 'Promoting access to public research data for scientific, economic, and social development', *Data Science Journal*, 3, pp.135–53.

Assante, M., Candela, L., Castelli, D. and Tani, A. (2016) 'Are scientific data repositories coping with research data publishing?', *Data Science Journal*, 15, 6.

Atici, L., Kansa, S., Lev-Tov, J. and Kansa, E. (2013) 'Other people's data: a demonstration of the imperative of publishing primary data', *Journal of Archaeological Method and Theory*, 20, 4, pp.663–81.

Austin, C., Brown, S., Fong, N. *et al.* (2015) 'Research data repositories: review of current features, gap analysis, and recommendations for minimum requirements' in *Proceedings of IASSIST Annual Conference*, Minneapolis, MN, June, pp.2–5.

Baker, S., Berger, J., Brady, P. *et al.* (2010) *Data-Enabled Science in the Mathematical and Physical Sciences: Technical Report*, National Science Foundation, Washington, DC.

Ball, A. (2014) *How to License Research Data*, Digital Curation Centre, Edinburgh.

Beagrie, Charles (2010) *Dryad Sustainability Plan: Interview Survey Findings*, available at http://wiki.datadryad.org/images/b/bf/Beagrie_suppdata_report_apr10.pdf (accessed May 2020).

Bicchieri, C. and Mercier, H. (2014) 'Norms and beliefs: how change occurs', in Xenitidou, M. and Edmonds, B. (eds) *The Complexity of Social Norms: Computational Social Sciences*, Springer, Cham, pp.37–54.

Bonneau, C., Kanthak, K., Berry, W. *et al.* (2015) 'Data access and research transparency: a joint statement by political science journal editors', *State Politics and Policy Quarterly*, 15, 4, 423–4.

Borgman, C. (2012) 'The conundrum of sharing research data', *Journal of the American Society for Information Science and Technology*, 63, 6, pp.1059–78.

Brown, C. (2003) 'The changing face of scientific discourse: analysis of genomic and proteomic database usage and acceptance', *Journal of the American Society for Information Science and Technology*, 54, 10, pp.926–38.

Callaghan, S. (2014) 'Preserving the integrity of the scientific record: data citation and linking', *Learned Publishing*, 27, 5, pp.15–24.

Campbell, J. (2015) 'Access to scientific data in the 21st century: rationale and illustrative usage rights review', *Data Science Journal*, 13, pp.203–30.

Crotty, D. (2016) 'What price progress? The costs of an effective data publishing policy', 13 January, available at https://scholarlykitchen.sspnet.org/2016/01/13/what-price-progress-the-costs-of-an-effective-data-publishing-policy/ (accessed May 2020).

Dallmeier-Tiessen, S., Darby, R., Gitmans, K. *et al.* (2012) 'Opportunities for data exchange', available at http://zenodo.org/record/8305#.Vw-IUfl94dU (accessed May 2020.

Dataverse (2017) *A Comparative Review of Various Data Repositories*, 25 July, available at https://dataverse.org/blog/comparative-review-various-data-repositories (accessed May 2020).

Eisenberg, R. (2006) 'Patents and data-sharing in public science', *Industrial and Corporate Change*, 15, 6, pp.1013–31.

Elias, P. and Entwisle, B. (2013) *New Data for Understanding the Human Condition: International Perspectives. Report on Data and Research Infrastructure for the Social Sciences*, available at www.oecd.org/sti/sci-tech/new-data-for-understanding-the-human-condition.pdf (accessed May 2020).

Eschenfelder, K. and Johnson, A. (2011) 'The limits of sharing: controlled data collections', *Proceedings of the American Society for Information Science and Technology*, 48, 1, pp.1–10.

Eschenfelder, K. and Johnson, A. (2014) 'Managing the data commons: controlled sharing of scholarly data', *Journal of the Association for Information Science and Technology*, 65, 9, pp.1757–74.

Fecher, B., Friesike, S. and Hebing, M. (2015) 'What drives academic data sharing?', *PloS One*, 10, 2, e0118053.

Harley, D., Acord, S., Earl-Novell, S., Lawrence, S. and King, C. (2010) *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines*, Center for Studies in Higher Education, University of California, Berkeley, CA.

Hrynaszkiewicz, I., Birukou, A., Astell, M. *et al.* (2017) 'Standardising and harmonising research data policy in scholarly publishing', *bioRxiv*, 4 April, available at www.biorxiv.org/content/biorxiv/early/2017/04/04/122929.full.pdf (accessed May 2020).

Hu, W-C. and Kaabouch, N. (2014) *Big Data Management, Technologies, and Applications*, IGI Global, Hershey, PA.

IFDO (2014) *Policies for Sharing Research Data in Social Sciences and Humanities*, available at www.cessda.eu/News-Events/News/CESSDA/IFDO-Report-2014-Policies-for-Sharing-Research-Data-in-Social-Sciences-and-Humanities (accessed May 2020).

Ioannidis, J., Allison, D., Ball, C. *et al.* (2009) 'Repeatability of published microarray gene expression analyses', *Nature Genetics*, 41, 2, pp.149–55.

Jasny, B. (2013) 'Realities of data sharing using the genome wars as case study – an historical perspective and commentary', *EPJ Data Science*, 2, 1, available at https://epjdatascience.springeropen.com/articles/10.1140/epjds13#citeas (accessed may 2020).

Kaye, J. (2012) 'The tension between data sharing and the protection of privacy in genomics research', *Annual Review of Genomics and Human Genetics*, 13, pp.415–31.

Kuipers, T. and van der Hoeven, J. (2009) *PARSE: Insight into Issues of Permanent Access to the Records of Science in Europe. Survey Report*, European Commission, Brussels.

Larivière, V., Haustein, S. and Mongeon, P. (2015) 'The oligopoly of academic publishers in the digital era', *PloS One*, 10, 6, e0127502.

McCain, K. (1995) 'Mandating sharing journal policies in the natural sciences', *Science Communication*, 16, 4, pp.403–31.

McCullough, B., McGeary, K. and Harrison, T. (2008) 'Do economics journal archives promote replicable research?', *Canadian Journal of Economics*, 41, 4, pp.1406–20.

Macfarlane, B. and Cheng, M. (2008) 'Communism, universalism and disinterestedness: re-examining contemporary support among academics for Merton's scientific norms', *Journal of Academic Ethics*, 6, 1, pp.67–78.

McSherry, C. (2009) *Who Owns Academic Work? Battling for Control of Intellectual Property*, Harvard University Press, Cambridge, MA.

Mauthner, N. and Parry, O. (2013) 'Open access digital data sharing: principles, policies and practices', *Social Epistemology*, 27, 1, pp.47–67.

Merton, R. (1973/1942) 'The normative structure of science' in Merton, R. (ed.) *The Sociology of Science: Theoretical and Empirical Investigations*, University of Chicago Press, Chicago, IL, pp.267–78.

Milia, N., Congiu, A., Anagnostou, P. *et al.* (2012) 'Mine, yours, ours? Sharing data on human genetic variation', *PloS One*, 7, 6, e37552.

Mongeon, P. and Paul-Hus, A. (2016) 'The journal coverage of Web of Science and Scopus: a comparative analysis', *Scientometrics*, 106, 1, pp.213–28.

Montgomery, K. and Oliver, A. (2009) 'Shifts in guidelines for ethical scientific conduct. How public and private organizations create and change. Norms of research integrity 1', *Social Studies of Science*, 39, 1, pp.137–55.

National Research Council (1997) *Bits of Power: Issues in Global Access to Scientific Data*, National Academies Press, Washington, DC.

Nelson, B. (2009) 'Data sharing: empty archives', *Nature*, 461, 7261, pp.160–3.

NISO (2013) *Recommended Practices for Online Supplemental Journal Article Materials*, available at www.niso.org/apps/group_public/download.php/10055/RP-15-(accessed May 2020).

PARSE Insight (2010) *Insight into Issues of Permanent Access to the Records of Science in Europe. Case Studies Report*, European Commission, Brussels.

Pienta, A., Alter, G., and Lyle, J. (2010) 'The enduring value of social science research: the use and reuse of primary research data', November, available at www.researchgate.net/publication/48927380_The_Enduring_Value_of_Social_Science_Research_The_Use_and_Reuse_of_Primary_Research_Data (accessed May 2010).

Piwowar, H. and Chapman, W. (2008) 'Identifying data sharing in biomedical literature', *AMIA Annual Symposium Proceedings*, pp.596–600.

Radder, H. (2010) 'Mertonian values, scientific norms, and the commodification of academic research' in Radder, H., *The Commodification of Academic Research: Analyses, Assessments, Alternatives*, University of Pittsburgh Press, Pittsburgh, PA, pp.231–58.

Ramin, S., and Sarraf Shirazi, A. (2012) 'Comparison between impact factor, SCImago journal rank indicator and Eigenfactor score of nuclear medicine journals', *Nuclear Medicine Review. Central and Eastern Europe*, 15, 2, pp.132–6.

Rappert, B. and Bezuidenhout, L. (2016) 'Data sharing in low-resourced research environments', *Prometheus*, 34, 3/4, pp.207–24.

Reichman, J. and Uhlir, P. (2003) 'A contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment', *Law and Contemporary Problems*, 66, 1/2, pp.314–462.

Savage, C. and Vickers, A. (2009) 'Empirical study of data sharing by authors publishing in *PLoS* journals', *PloS One*, 4, 9, e7078.

Stephens, Z., Lee, S., Faghri, F. *et al.* (2015) 'Big data: astronomical or genomical?', *PLoS Biol*, 13, 7, e1002195.

Sturges, P., Bamkin, M., Anders, J. *et al.* (2015) 'Research data sharing: developing a stakeholder-driven model for journal policies', *Journal of the Association for Information Science and Technology*, 66, 12, pp.2445–55.

Swan, A. and Brown, S. (2008) *To Share or Not to Share: Publication and Quality Assurance of Research Data Outputs. Technical Report*, Research Information Network, London.

Tenopir, C., Allard, S., Douglass, K. *et al.* (2011) 'Data sharing by scientists: practices and perceptions', *PloS One*, 6, 6, e21101.

Tenopir, C., Dalton, E., Allard, S. *et al.* (2015) 'Changes in data sharing and data reuse practices and perceptions among scientists worldwide', *PloS One*, 10, 8, e0134826.

Unsworth, J. (2006) *Our Cultural Commonwealth: The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*, ACLS, New York.

Vines, T. (2017) 'Is there a business case for open data?', *Scholarly Kitchen*, 15 November, available at https://scholarlykitchen.sspnet.org/2017/11/15/business-case-open-data/ (accessed May 2020).

Vines, T. Andrew, R. Bock, D. *et al.* (2013) 'Mandated data archiving greatly improves access to research data', *FASEB Journal*, 27, 4, pp.1304–8.

Vlaeminck, S. (2013) 'Data management in scholarly journals and possible roles for libraries – some insights from EDaWaX', *Liber Quarterly*, 23, 1.

Weller, T. and Monroe-Gulick, A. (2014) 'Understanding methodological and disciplinary differences in the data practices of academic researchers', *Library Hi Tech*, 32, 3, pp.467–82.

Womack, R. (2015) 'Research data in core journals in biology, chemistry, mathematics, and physics', *PLoS One*, 10, 12, e0143460.

Womack, R. (2015) 'Research data in core journals in biology, chemistry, mathematics, and physics', *PLoS One*, 10, 12, e0143460.

Zachariah, R., Kumar, A., Reid, A. *et al.* (2014) 'Open access for operational research publications from low and middle-income countries: who pays?', *Public Health Action*, 4, 3, pp.141–4.

Zenk-Möltgen, W. and Lepthien, G. (2014) 'Data sharing in sociology journals', *Online Information Review*, 38, 6, pp.709–22.