

Open-access policy and data-sharing practice in UK academia

Yimei Zhu

University of Leicester, UK

Abstract

Data sharing can be defined as the release of research data that can be used by others. With the recent open-science movement, there has been a call for free access to data, tools and methods in academia. In recent years, subject-based and institutional repositories and data centres have emerged along with online publishing. Many scientific records, including published articles and data, have been made available via new platforms. In the United Kingdom, most major research funders had a data policy and require researchers to include a ‘data-sharing plan’ when applying for funding. However, there are a number of barriers to the full-scale adoption of data sharing. Those barriers are not only technical, but also psychological and social. A survey was conducted with over 1800 UK-based academics to explore the extent of support of data sharing and the characteristics and factors associated with data-sharing practice. It found that while most academics recognised the importance of sharing research data, most of them had never shared or reused research data. There were differences in the extent of data sharing between different gender, academic disciplines, age and seniority. It also found that the awareness of Research Council UK’s (RCUK) Open-Access (OA) policy, experience of Gold and Green OA publishing, attitudes towards the importance of data sharing and experience of using secondary data were associated with the practice of data sharing. A small group of researchers used social media such as Twitter, blogs and Facebook to promote the research data they had shared online. Our findings contribute to the knowledge and understanding of open science and offer recommendations to academic institutions, journals and funding agencies.

Keywords

Data sharing; open access; open science; scholarly communication; science communication; social media

1. Introduction

New forms of scholarly communication emerged with the development of the Internet and digital technology which enabled more open practices within the academic community. These new forms include open-access (OA) publishing, sharing primary research data, publishing research updates online as well as using social media for various reasons in research work, such as searching for research information, promoting publications and networking with peers. The Open Science Movement emerged in this new age. Open science not only supports the concept of OA publishing of research articles but also extends the open practice to publishing datasets, workflow, methods, details of ongoing research processes and so on [1].

In academia, OA policy has been discussed among various academic communities and a number of research councils in the United Kingdom have had policies on OA since 2005. From April 2013, Research Council UK’s¹ (RCUK) OA policy came into effect which required RCUK-funded research to be published either through OA journals (Gold OA) or self-archiving (Green OA) [2]. Along with the OA movement for publishing research articles, a growing chorus of voices have advocated free access to research data, tools and methods [3]. In recent years, subject-based and institutional repositories, data centres, open-source software and public copyright licences have emerged along with online publishing. Many scientific records, including published articles, data, presentation slides, methods and tools have been made freely available to all Internet users via those new platforms [4,5]. The rationales and purposes of OA to research data are related to the accessibility to public goods, validation of findings, reducing duplicate data collection and accelerating scientific progress [6]. The Berlin Declaration on Open Access to Knowledge in the Science and Humanities (2003)

Corresponding author:

Yimei Zhu, School of Media, Communication and Sociology, University of Leicester, LE1 7JA, UK.

Email: yz411@le.ac.uk

supported an OA paradigm and called for the setting up of regulation to promote OA to scientific contributions, including 'original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material'. In the United Kingdom, most major UK research funders have a data policy and require researchers to include a 'data-sharing plan' when applying for funding [7].

However, there are a number of barriers to the full-scale adoption of data sharing. Those barriers are not only technical (as infrastructure needs to be built to store the data), but also psychological and social. The lack of recognition incentive for sharing research data, the lack of standard for data formats and metadata, privacy issues, fear of misuse, lack of time and resources are all potential barriers that can hinder data sharing [8,9]. Some disciplines such as Biomedical Sciences and Environmental Sciences have a long history of a data-sharing culture, while others may not produce primary research data. Academics might be more likely to share data if related articles were published in an OA journal or a journal with a stronger data-sharing policy or if they had prior experience sharing or reusing data [10].

In order to examine the attitudes and practice towards data sharing and their relations with OA policy, a survey was conducted with over 1800 UK-based academics. This study aims to explore the extent of support of data sharing across all disciplines and the characteristics and factors associated with the practice of data sharing.

2. Background and context

Borgman [11] defined the term 'data sharing' as releasing research data that can be used by others. The releasing of data can take many forms, including private exchange upon request and deposit in an online repository. Wallis et al. [12] suggested that 'data' are the information that can be reinterpreted. Soranno et al. [13] defined data sharing as depositing data in publicly accessible institutional repositories at publication or at the end of a project which is a narrower definition of data sharing. While acknowledging that free and OA to research data would encourage citizen scientists and wider participation, data sharing should not be restricted to sharing data in an OA repository. Eschenfelder and Johnson [14] suggested that open-data archives could provide OA to anonymous public use including for commercial purposes, but also could have controlled access which require the users to register and provide information as a condition of access or use. There are also data journals that publish metadata documents describing a particular data set or a group of data sets which are online and accessible for sharing [15].

Borgman [11] suggested that rationales of sharing data are to reproduce or verify research, to make the results of publicly funded research available to the public, to enable others to ask new questions of extant data and to advance the state of research and innovation. However, data sharing can be limited between strangers or through online data repositories. Research by Nelson [16] suggests that very few academics share their data with other scientists or the public more widely. Kennan et al. [17] interviewed active members of a large environmental voluntary group who had data-sharing experiences. They found that data sharing within the group occurred frequently, but restricted among oral networks and very rarely beyond their members and associated groups through online repositories. Hannay [18] suggests that the barriers to the full-scale adoption of data sharing are not only technical (as infrastructure needs to be built to store the data), but also psychological and social. Their main concern of data sharing in Kennan and colleagues' study is trust. There was the lack of recognition incentive for sharing research [8]. Moreover, the generosity of sharing one's own research data might allow credit to be inappropriately allocated to reusers [7]. An anecdotal example of this is provided by Cragin et al. [19], who report that a reuser of data from a geological study published his findings first, while the original author was still working on his analysis. In that case, the original researcher who collected the data lost his 'priority' in that specific scientific discovery. Junior academics, who have not yet secured a professional position or reputation, particularly fear having data or findings, which can generate multiple reports over time, taken by others, as the academic reward system is associated with publication and its impact [20].

It is notable that although the principles for open data sharing among the genomics community were agreed in Bermuda in 1996, in practice, it was difficult to convince researchers to follow the principles. Data policies are more established in some subject areas than others, such as in Biomedical Sciences. Many journals in this area have started to request the sharing of primary datasets from authors. In a study of journal policies and data sharing, Piwowar and Chapman [21] studied biomedical journals ($n = 70$) that published articles on 'gene expression profiling' in 2006 and analysed journal data-sharing policies applicable to microarrays. They found that academic publishers were more likely to have stronger data-sharing policies than commercial publishers, while all four of the OA journals in the sample had a data-sharing policy. Policy strength was associated with impact factor – journals with a higher impact factor were more likely to have stronger data-sharing policy. Policy strength was also positively correlated with measured data-sharing submission into the GEO database, in that the journals with a stronger policy had higher median data-sharing prevalence than journals with no data-sharing policy. However, data-sharing prevalence in general was quite low, even for journals with very strict sharing requirements (29%).

In the United States, the National Institutes of Health (NIH) have a data-sharing policy; however, being subject to the NIH data-sharing mandate requirement are not necessarily more likely to share their own datasets. Piwowar and Chapman [22] analysed 397 biomedical microarray articles and found that researchers were more likely to practise data sharing when their work was published in a high-impact journal and when the first or last authors were highly experienced in their careers with a high level of professional impact. In another study, Piwowar [10] used similar methods and found that authors were more likely to share data if related articles were published in an OA journal or a journal with a stronger data-sharing policy, if they had prior experience sharing or reusing data, or if their research received a large amount of NIH funding.

The disciplinary difference of data sharing was evident in a few studies. Soranno et al. [13] suggested that disciplines such as meteorology, astronomy and genomics have shown an emerging norm of data sharing in relation to substantial benefits to those disciplines and to the society, but not so much in environmental science disciplines. Disciplinary culture may influence the practice of sharing research data. Some disciplines may have a long history of a data-sharing culture, while others may not produce primary research data. In those disciplines with more established data policies such as Biomedical Sciences, Physical Sciences and Social Sciences, researchers may have more resources and skill sets to store their primary research data (e.g. the availability of repositories and standardisation of datasets). In the Biomedical Sciences, journals often have data-sharing policies especially those OA journals, and hence, authors are more likely to deposit research data when they publish articles to comply with the journals' requirement [10]. In Humanities, some disciplines do not produce research data that can be reused by others as discussed earlier. However, a study found that a key challenge for Humanities scholars is their lack of ability with tools and methods to link data housed in different archives [7]. A recent study also identified infrastructure issues as one of the main barriers for researchers in Humanities to share data because there is no single common practice across Humanities, thus calling for the development of cross-domain infrastructure components to allow data sharing among Humanities [23]. Ethical issues involve human subjects may affect academics' willingness and ability to share research data. Disciplines in Medical and Social Sciences with human participants may be reluctant to share research data because of confidentiality or anonymity reasons [9]. Natural Sciences and Engineering have no human participants and thus would not face such issues.

Kim and Burns [24] collected responses from 608 biological scientists based in the United States found that pressures by funding agencies and journals and availabilities of data repositories and metadata standards have positive influence on data-sharing behaviours. They also found that positive attitudes towards data sharing could predict data-sharing behaviour. His study also regarded data sharing not necessarily with full free access. However, they excluded research students in their final analysis. Another study by Kim and Stanton [25] found that data-sharing behaviours are positively associated with regulative pressure by journals, normative pressure at a discipline level and scholarly altruism at an individual level. Tenopir et al. [26] conducted a survey with international academics on two occasions in 2009 ($n = 1329$) and 2013 ($n = 1015$). They found that the willingness to share data and the self-reported data-sharing behaviour had increased. Young academics felt more favourably towards data sharing and reuse, yet making less of their data available to share than older academics. The results also indicated different constraints and enablers of data sharing and reuse across different subject disciplines. However, the authors failed to inform the participants what data is and the survey study assumed every participant produced research data. Thus, the results may have included academics who did not produce primary research data that are reusable. Another limitation is that there were very few numbers of respondents in Arts and Humanities.

With the development of new technology, academics have been discussing science and sharing research work on social media sites [27]. Social media tools such as blogs and Twitter have been adopted by academics and scientists to record and disseminate research work and facilitate debates among colleagues from various parts of the world [28,29]. Social media tools may be used to promote publications and increase readership [30]. Twitter was found being effective in promoting publication. A study found that Twitter citations were significantly faster than citations in traditional media, as 39% of citations in their sample referred to articles less than 1 week old, and 15% of them referred to papers published that same day [31]. Hence, social media can also be adopted by individual researcher to promote the sharing of their research data. To date, there has not been any study examining the relationship between primary data sharing and social media use.

Having reviewed relevant literature, it is notable that most studies on research data sharing were conducted in the United States. Little was known on how UK academics supported data sharing and what the benefits and barriers were. Existing literature indicated a difference between groups of academic discipline, age and seniority in data-sharing behaviour. Although gender difference in data-sharing behaviour was not evident in existing studies, some studies found that men were more likely to accept the use of new technology than women [32,33]. Men were also found to be more likely to have experience of OA publishing than women [34]. Hence, this study would examine the differences in data-sharing between male and female. Factors such as social media use, experience of OA publishing, awareness of OA policy,

attitudes towards the importance of data sharing and experience of using secondary data will also be examined with the likelihood of data-sharing behaviour. Hence, four research questions emerge:

RQ1. To what extent do UK academics support data sharing?

RQ2. What are the rationales and barriers towards data sharing for UK academics?

RQ3. Were there differences in data sharing by gender, academic discipline, age and seniority?

RQ4. What other factors (social media use, OA publishing, awareness of OA policy, attitudes towards the importance of data sharing and experience of using secondary data) are associated with academics' data-sharing behaviour?

3. Methods

An online survey was conducted with academics based in the United Kingdom with regards to their scholarly communication practice.² The sampling frame of the survey consisted of the population of all academics based in the Russell Group universities (24 in total). Russell Group universities all have a strong research focus. As such the research would be capturing the attitudes and behaviour of academics at leading research-focused universities. Furthermore, Russell Group universities are large universities with a broad range of disciplines for both Science and Humanities. In all, 12 Russell Group universities were randomly selected and an email invitation was sent to academics based in those universities using publicly available information on their university websites. The survey was conducted at summer 2013. In total, 1829 valid responses were received with a response rate at 4.4%. One section of the survey questionnaire asked respondents' attitudes and experiences in data sharing. Respondents were asked whether they had deposited primary research data in an online repository that can be reused by other researchers. Furthermore, they were asked whether they had used secondary research data that were collected by others. Respondents were required to rate the importance of making research data available online for reuse. In addition, respondents were questioned whether they were aware of RCUK's OA policy which came into effect 3 months before the survey was distributed and they were given the choices of 'yes', 'no', 'heard of, but not sure about the detail' and 'other (please specify)' to the question. Respondents were asked whether they had published research articles in OA journals and whether they had deposited research articles in OA repositories. They were asked whether they used a number of social media tools (e.g. Twitter, Facebook and research blogs) for their research work. Respondents were also given options to leave comments for a number of individual questions with regard to experiences and attitudes towards data sharing. For example, one of the answer options was 'other or/and comments' for questions related to data-sharing experience. The attitudes towards data-sharing question included a follow-up enquiry 'any reason of this (please specify)' with space for comments.

Qualitative data (comments provided by respondents in relation to data-sharing attitudes and behaviour) were coded and analysed using thematic analysis in order to answer RQ2. Thematic analysis is a method for identifying, analysing and interpreting patterns and themes of qualitative data in rich detail [35].

Quantitative survey data were coded and analysed in SPSS. Descriptive analysis was conducted to answer RQ1 and RQ3. Cross tabulations were applied to measure the associations between academics' experience of data sharing with discipline areas, gender, age and seniority. A cross tabulation is a useful procedure or technique for measuring the association between two variables [36]. Chi-square test was applied to examine the significance level of these associations. In addition, logistic regression modelling was applied to investigate for RQ4. Logistic regression methods are useful in models with a binary dependent variable for determining whether or not a particular independent variable has a statistically significant association with the outcome variable, having allowed for other independent variables in the model [37]. This modelling method is based on the notion of probabilities – the probability of an event occurring over the probability of it not occurring. In this study, it explores whether other factors are associated with the likelihood of academics sharing their own research data.

Valid responses consist of 46% female and 54% male. The disciplines include Medical and Life Sciences (35%), Natural Sciences and Engineering (23%), Social Sciences (27%) and Arts and Humanities (15%). Respondents include 35% aged under 35 years, 26% aged 35–44 years, 21% aged 45–54 years and 18% aged 55 and over. Seniority was categorised as researchers in training (20%) including PhD candidates, master students and research assistants, lecturers/ research fellows/post-docs (39%), senior lecturers/senior researchers (16%) and professors/readers (24%).

4. Findings

4.1. Data-sharing behaviour and attitudes

Table 1 indicates that the vast majority of respondents acknowledged the importance with 1459 out of 1695 (86%) respondents rating it 'very important' or 'fairly important' to make research data available online for reuse. Around

Table 1. Attitudes towards the importance of making research data available online for reuse.

	<i>N</i>	%
<i>How important do you think it is, in general, to make research data available online for reuse?</i>		
Very important	673	40%
Fairly important	786	46%
Not very important	211	12%
Not at all important	25	1%
Total	1695	100%

Table 2. Experience of depositing primary data in online repositories.

	<i>N</i>	%
<i>Have you deposited your own primary research data in an online repository that can be reused by other researchers?</i>		
Yes	360	21%
No	1364	79%
Total	1724	100%

Table 3. Experience of using secondary data from online repositories.

	<i>N</i>	%
<i>In your research work, have you used secondary research data from an online repository that were collected by other researchers?</i>		
Yes	499	29%
No	1229	71%
Total	1728	100%

Table 4. Experience of promoting shared primary research data.

	<i>N</i>	%
<i>If yes, did you do anything to promote the most RECENT primary research data that you deposited online for reuse? (Tick all that apply)</i>		
No	236	68%
Yes, I told colleagues in person.	77	22%
Yes, I emailed the information about data to colleagues.	66	19%
Yes, I posted info and link on Twitter/research blogs/Facebook.	27	8%
Total	346	100%

one-fifth of respondents had experience of sharing their own data. As shown in Table 2, of 1724 respondents, 360 (21%) reported having deposited their own primary research data in online repositories that could be reused by other researchers. A slightly higher percentage of respondents answered 'yes' (29%, 499 out of 1728) to having used secondary research data collected by others in their research work, as shown in Table 3.

As shown in Table 4, among the 346 respondents who had shared data, only 8% indicated having promoted their data via social media tools of Twitter, Facebook or research blogs. Some academics informed colleagues about their data in person (22%) or through emails (19%). The majority who deposited their data in an online repository that can be used by others did nothing (68%) to promote it.

4.2. Data-sharing rationales and barriers

The rationales and barriers suggested by the respondents of this study were consistent with findings from previous studies. Many respondents indicated that publicly funded research data should be accessible for the public, and many

acknowledged the benefits such as improving quality, reducing instances of fraud, validating research findings, avoiding duplicate data collection, saving time and cost, maximising the use of resources and encouraging collaboration.

One of the most common rationales of making data available online for reuse was for the public good as commented by a researcher in Physics:

Data that is produced through publically funded research should be available for all. (Male, Research fellow/post-doc, Physics)

Many respondents acknowledged the benefits of having primary research data available online for reuse, such as improving quality and reducing instances of fraud. As one lecturer in Economics and Econometrics stated,

[Data sharing] faster progress of research if data is public and allow verifying the results. (Female, Lecturer, Economics and Econometrics)

Making primary research data available online for reuse can advance science by validating research findings, avoiding duplicate data collection, saving time and cost, maximising the use of resources and encouraging collaboration. These advantages were highlighted by a number of respondents from various discipline areas:

I think that secondary analysis is becoming increasingly important. Also, a lot of primary data is collected, perhaps analysed once or twice for thesis, a report or a paper and then is forgotten about—which is a waste! Finally, it is very useful for research data to be available online for comparative purposes. (Female, Lecturer, Education)

It reduces the resources required (time and money) to make research discoveries and encourages collaboration. (Female, Lecturer, Psychology, Psychiatry and Neuroscience)

The requirements from major UK funding councils and some journals such as PLOS One were also taken into account as a reason why some academics shared their primary data.

A number of challenges of data sharing were identified as common themes, such as ethical issues, risks of losing priority to competitors, uncertainty of the usefulness of their research data to others and excessive time and effort needed for producing reusable data. One of the challenges of data sharing were related to what kind of data researchers produced and the time and effort required to produce detailed metadata alongside the primary data; otherwise, the data could be misinterpreted by others without firsthand insights. This issue could be a barrier in Sciences or Humanities as stated by respondents from various discipline areas:

It's just very time-consuming to do so and provide the necessary documentation. I make my code available but not usually my data. (Female, Reader, Psychology, Psychiatry and Neuroscience)

It depends on the nature of the data, and outside the framework of the project it was gathered for it may be less useful, or may possibly be misinterpreted. (Female, Research fellow/post-doc, History)

I'm happy to collaborate but the use of data without insight has a major risk for misinterpretation. (Female, Senior lecturer, Clinical Medicine)

There were also concerns related to ethical and confidentiality issues, especially when there were human subjects involved. For example, a lecturer in Anthropology stated the challenge of privacy and confidentiality of human participants in their research:

I am an anthropologist and my field notes are deeply private collections, and I would worry about putting it online for fear of risking informant confidentiality, and safety. Also it's hard to see how different people could use them. (Male, Lecturer, Anthropology)

Similarly, a PhD student in History raised an ethical and privacy issue:

It depends on the type of data. I can see instances where it would be very important (maybe in science when repeat experiments could be done) but this is not always the case. Some of my interviews I would be very reluctant to put online for ethical/privacy reasons. (Female, PhD candidate, History)

The respondents who selected 'no' to whether they had deposited their primary research data online were asked whether they would share their 'primary research data in an online repository that can be reused by other'. Among the 1303 respondents who had not deposited primary research data in online repositories, 10% (135) selected 'no, because of ethical issues' when asked whether they would share primary research data online to be reused by others in the future.

Respondents from Medical and Life Sciences and Social Sciences and Humanities were more likely to be concerned about ethical issues compared with respondents from Natural Sciences and Engineering. For the same question, 10% (126 out of 1303) of respondents who had not deposited data in online repositories selected 'no, I want to secure publication'. Another 39% (512 out of 1303) stated 'not sure'.

It seems academic competition can be a major barrier for sharing data as academics often depend on their primary data to publish findings and advance their career. A lecturer from Biological Sciences highlighted concerns about competition:

Important for genetic data, but I am not sure that all data needs to be available to everyone immediately because of issues with competition. (Female, Lecturer, Biological Sciences)

Some academics indicated that they only wanted to share with colleagues privately and with whom they could collaborate to produce research papers together. For example, a research fellow/post-doc in Sociology explained her reason of choosing 'not very important' for making research data available online for reuse:

I think researchers should freely share research data with those colleagues who email and ask personally. So if someone is interested in the primary data that I have collected then they can email me, I will share the data and we will produce a paper together. I resent, however, spending the time and effort gaining ethics approval, securing access to respondents, spending the time collecting the data and then transcribing all of the interviews for someone else to swoop in and just use the end product (i.e. the data that I produced). (Female, Research fellow/post-doc, Sociology)

A researcher could spend considerable time preparing the primary data they have collected. Sharing would be potentially benefitting competitors. A number of respondents commented that they were unsure about the benefits and usefulness of sharing primary research data and the usefulness of sharing may vary for different disciplines. For example, a research fellow/post-doc in Engineering suggested disciplinary differences in primary research data's usefulness:

There are existing repositories that work well for data intensive research (genomics, proteomics), but not all research really benefits from these. Most of my work is deposited in institution data centres and can be accessed upon enquiry, but it is also within the publication (in the shape of statistical summaries) and the usefulness of the raw data is unclear. (Male, Research fellow/post-doc, General Engineering)

4.3. Differences in data sharing by gender, academic discipline, age and seniority

Table 5 summarises the descriptive analysis results of data-sharing experience by gender, academic discipline, age and job grade (seniority). All the results in Table 5 are statistically significant, except for discipline. The results suggest that there were significant differences between gender, age and seniority for sharing primary research data. In general, men (24%) were more likely to share primary research data online compared with women (17%). Respondents who were older and in senior job grades had more experience sharing research data. Those aged under 35 years (13%) were the least likely to have shared research data compared with 26% aged 45 years and over and 23% aged 35–44 years. Professors/readers (35%) were more likely to have deposited research data for reuse compared with 9% of researchers in training. Respondents in Natural Sciences and Engineering seemed to be slightly more likely to share primary research data online, but the differences were small and not significant.

4.4. Factors associated with the likelihood of data-sharing behaviour

Logistic regression models were fitted to explore what factors were associated with the likelihood of academics sharing primary research data. Logistic regression analysis was run in SPSS using a stepwise procedure, which only included gender, discipline and age in the first block. Because age and job grade in this study were found to be highly correlated with one another, in order to avoid multicollinearity, only age was included in as an independent variable in the logistic regression analysis. The second block added social media experience, awareness of RCUK's OA policy and experience of publishing in OA journals. The third block added attitudes towards the importance of data sharing. The fourth and final blocks added experience of depositing research articles in online repositories and experience of using secondary data. The stepwise method enables the modelling results to compare significance levels of independent variables after including other independent variables. Table 6 reports the coefficient *B* values and their odds ratios (ORs).

Table 5. Experience of data sharing by gender, academic discipline, age and seniority.

		<i>Have you deposited your own primary research data in an online repository that can be reused by other researchers?</i>			
		Yes		No	
		<i>N</i>	%	<i>N</i>	%
Gender	Female	132	17%	651	83%
	Male	226	24%	701	76%
	<i>p</i> value	0.000***			
Academic discipline	Medical & Life Sciences	124	21%	479	79%
	Natural Sciences & Engineering	99	26%	289	74%
	Social Sciences	87	19%	376	81%
	Arts & Humanities	49	19%	213	81%
	<i>p</i> value	0.070			
Age (years)	Under 35	73	13%	507	87%
	35–44	103	23%	340	77%
	45–54	99	26%	280	74%
	55 and over	83	27%	230	73%
	<i>p</i> value	0.000***			
Seniority	Researchers in training	27	9%	287	91%
	Lecturers/research fellows/post-docs	128	19%	529	81%
	Senior lecturers/senior researchers	58	21%	213	79%
	Professors/readers	140	35%	264	65%
	<i>p</i> value	0.000***			

The result of Model 1 indicates that having allowed for gender and age in the model, academics in Natural Sciences and Engineering were more likely than those in Medical and Life Sciences ($p < 0.05$) and those in Arts and Humanities ($p < 0.05$) to share research data. However, no significant difference was found for academics in Social Sciences compared with being in Natural Sciences and Engineering.

Model 2 results indicate that having controlled for demographic factors, those who used at least one of the three social media tools (Twitter, Facebook or research blogs) for their research work were more likely to share data. Those who were aware of RCUK's OA policy were more likely than those who said 'no' ($p < 0.001$) and who said 'heard of, but not sure about the detail' ($p < 0.01$) to share data. Those that had published in Gold OA journals were more likely to share data than those who had not used Gold OA publishing. Models 3 and 4 added attitudes towards the importance of data sharing and experience of depositing articles in online repositories. Perhaps not surprisingly respondents who rated it 'very important' and 'fairly important' were much more likely than those who rated it 'not important' to share data. Having allowed for attitudes towards the importance of data sharing, there was no significant difference between those who had used social media or not in terms of sharing primary research data. Moreover, being in Arts and Humanities had no significant difference from being in Natural Sciences and Engineering after taking account of attitudes towards the importance of data sharing. The result of Model 4 indicates that academics who had deposited their research articles in online repositories (Green OA) were more likely to deposit research data. Gold OA publishing experience was no longer significant after taking into account Green OA experience.

The result of Model 5 indicates that experiences of using secondary data from online repositories are positively associated with experience of sharing research data. Age, attitude towards the importance of sharing data and Green OA experience were still significant after controlling for experience of using secondary data. Answering 'no' for the awareness of RCUK's OA policy was no longer significant. This was perhaps because awareness of RCUK's OA policy was associated with Green OA experience [34].

Hence, positive attitudes towards the importance of making research data available online for reuse were positively associated with the likelihood of sharing primary research data in online repositories. Experiences of Green OA publishing and using secondary data were positively associated with the likelihood of sharing data. This suggests that those who had used repositories to deposit articles might be more likely to deposit data alongside their articles as part of funding requirements and have become more skilled as regard how to use repositories to deposit data. Experience of using social media had some association with data-sharing experience. Awareness of RCUK's OA policy and Gold OA experience had some associations with data-sharing experience. Academics who stated that they had shared data online were more likely to have used secondary data and deposited research articles in online repositories, thus more likely perhaps to have learnt RCUK OA policy. For academics who had no experience of using secondary data, being aware of the RCUK OA

Table 6. Logistic regression analysis on the likelihood of sharing primary research data.

		Model 1		Model 2		Model 3		Model 4		Model 5	
		B	OR	B	OR	B	OR	B	OR	B	OR
Background characteristics	Gender (reference: female)	0.32	1.37*	0.28	1.32*	0.23	1.25	0.24	1.27	0.14	1.15
	Discipline (reference: Natural Sciences & Engineering)										
	Medical & Life Sciences	- 0.41	0.66*	- 0.47	0.62**	- 0.43	0.65*	- 0.28	0.76	- 0.22	0.80
	Social Sciences	- 0.35	0.71	- 0.25	0.78	- 0.13	0.87	- 0.03	0.97	- 0.10	0.90
	Arts & Humanities	- 0.55	0.60*	- 0.47	0.62*	- 0.38	0.68	- 0.28	0.76	- 0.27	0.76
	Age (years; reference: under 35)										
35-44	0.81	2.26***	0.67	1.94***	0.65	1.92***	0.57	1.77**	0.60	1.82**	
45-54	0.89	2.43***	0.74	2.09***	0.80	2.22***	0.67	1.96**	0.71	2.04**	
55 and over	0.94	2.57***	0.83	2.30***	0.84	2.31***	0.78	2.19***	0.91	2.49***	
Social media experience	Use of social media in research work (reference: no)			0.35	1.42**	0.17	1.19	0.09	1.10	- 0.09	0.92
Awareness of RCUK OA policy	Awareness of RCUK policy (reference: yes)										
	No			- 0.83	0.44***	- 0.77	0.46***	- 0.53	0.59**	- 0.39	0.67
	Heard of, but not sure about the detail			- 0.43	0.65**	- 0.45	0.64**	- 0.31	0.73	- 0.16	0.85
Gold OA experience	Gold OA publishing experience (reference: no)			0.45	1.57**	0.42	1.53**	0.25	1.28	0.30	1.35
Attitudes towards the importance of sharing data	OA importance (reference-not important)										
	Very important					2.41	11.15***	2.36	10.57***	1.61	4.98***
	Fairly important					1.14	3.12***	1.12	3.06***	0.73	2.07*
Green OA experience	Green OA publishing experience (reference: no)						1.02	2.78***	1.02	2.78***	
Use of secondary data	Having used secondary data (reference: no)									1.83	6.25***
	Constant	- 1.80	0.17***	- 1.76	0.17***	- 3.37	0.03***	- 3.88	0.02***	- 4.08	0.02***
Nagelkerke R ²		0.05		0.10		0.22		0.27		0.38	
N = 1471											

B: coefficient; OR: odds ratio; OA: open access; RCUK: Research Council UK.

Significance level of OR: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

policy had no association with sharing primary research data. Gold OA publishing experience was no longer significant having controlled for Green OA experience. This suggests that for those who had no experience of depositing research articles online, there is no significant difference from those who published in OA journals or not in terms of data sharing.

5. Discussion and implication

Sharing primary research data may improve accessibility to publicly funded research data, advance scientific progress and encourage collaboration. Many academics acknowledged the rationale of sharing research data and recognised the goals such as providing accessibility of publicly funded data to the public, validating research findings and avoiding duplicate data collection. However, many academics seemed to prioritise securing academic rewards and sharing data might work against that priority. Other barriers identified included the time and effort required to produce detailed meta- data alongside the primary data, ethical and confidentiality issues and uncertainty of the benefits and usefulness of sharing primary data.

Positive attitudes towards the importance of data sharing are associated with sharing activities. In relation to the activities of using secondary data, re-using data collected by others might influence academics to be reciprocal and share their own research data. Although it was also possible that these academics generally liked to share and they might come across others' data after they shared their own on repositories. Academics with experience of depositing research articles were more likely to deposit data, which might be part of funding requirements or being more skilled in how to use repositories to deposit research data. The differences in data-sharing experiences between those who published in OA journals and those who did not, as suggested by Piwovar [10], could be explained by experience with self-archiving. Academic institutions with their own OA repositories need to promote their service to be known and understood by more academics if they want their service to be fully used. Institutions without OA repositories need to find alternative ways to help their academic staff comply with UK Research and Innovation and HEFCE's OA policies in the future.

The academic community could benefit from OA to research data to validate findings and accelerate scientific progress. However, barriers such as lack of incentives and standards could prevent academics from sharing. This is especially the case for younger and junior academics who are in greater need of securing publication and funding to advance their career while sharing primary research data might jeopardise their chances of publishing before competitors. An incentive system and approach to the citing of data and databases are needed to promote data sharing in the future. Developing and adopting widely recognised, usable technological and descriptive standards can avoid the need for individual database-to-database negotiation and can lower the barriers of technical difficulties such as format differences, while diverse models of data sharing and collaborations should be encouraged and supported by academic institutions, journals and funding agencies [20]. Standards should be developed for sub-disciplines so that scientists have common and uniform data-sharing practices in their own research community [19]. Funding agencies, academic journals and institutions should also develop approaches to disseminate and promote data-sharing policies and standards, as many researchers were unaware of existing policies and standards. Knowing the related policies and standards early in the research cycle can help researchers identify sharable information and facilitate datasets in an appropriate format for deposition [19]. Academic institutions should reinforce these OA policies to academics through various channels including training especially for those at the start of their careers such as PhD students and research assistants who are more likely to be unaware of those policies. Funding agencies and academic institutions should also fund and maintain infrastructure for data sharing, including providing training and support for researchers who intend to share data [8]. This is especially important for those disciplinary areas without cross-domain infrastructure or common practice [23].

Academics would also be more willing to share primary research data if made aware of the potential benefits such as increased citation impact. Social media could be adopted to help academics to promote their publication and shared research data and potentially increases readership and citation [31]. Academic institutions could organise more social media training to help those who want to improve their digital skills for communicating and sharing their research outputs and data. Data policies are more established in some subject areas than the others and studies on developing strategies to encourage data sharing mainly focused on biomedical areas. As the format and volume of research data vary largely between and within disciplines, further studies of data sharing need to focus on individual discipline areas and especially those that have not been studied systematically.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Notes

1. Research Council UK is now called 'UK Research and Innovation' as of June 2018.
2. A full survey questionnaire can be accessed at: <https://yimeizhuereasearch.files.wordpress.com/2018/11/final-survey-questionnaire.pdf>

References

- [1] Grand A, Bultitude K, Wilkinson C et al. Muddying the waters or clearing the stream? Open science as a communication medium. In: *11th international conference on public communication of science and technology*, New Delhi, India, 6–9 December 2010, pp. 1–7. New Delhi: INPCST.
 - [2] RCUK. RCUK policy on open access, <http://www.rcuk.ac.uk/research/openaccess/> (2013, accessed 21 November 2014).
 - [3] Nicholson SW and Bennett TB. Data sharing: academic libraries and the scholarly enterprise. *Portal: Libr Acad* 2011; 11: 505–516.
 - [4] Björk BC. Open access to scientific publications - an analysis of the barriers to change? *Inform Res* 2004; 9: 170.
 - [5] De Roure D, Goble C, Aleksejevs S et al. Towards open science: the myExperiment approach. *Concurr Comp: Pract E* 2010; 22: 2335–2353.
 - [6] Piwowar HA, Day RS and Fridsma DB. Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2007; 2: e308.
 - [7] RIN. Data centres: their use, value and impact, www.rin.ac.uk/system/files/attachments/Data_Centres_Report.pdf (2011, accessed 17 February 2016).
 - [8] Piwowar HA, Becich MJ, Bilofsky H et al. Towards a data sharing culture: recommendations for leadership from academic health centers. *PLoS Med* 2008; 5: 1315–1319.
 - [9] Anderson N and Edwards K. Building a chain of trust: using policy and practice to enhance trustworthy clinical data discovery and sharing. In: *Proceedings of the 2010 workshop on governance of technology, information and policies*, Austin, TX, 7 December 2010, pp. 15–20. New York: ACM.
 - [10] Piwowar HA. Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLoS ONE* 2011; 6: e18657.
 - [11] Borgman CL. The conundrum of sharing research data. *J Assoc Inf Sci Tech* 2012; 63: 1059–1078.
 - [12] Wallis JC, Rolando E and Borgman CL. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS ONE* 2013; 8: e67332.
 - [13] Soranno PA, Cheruvilil KS, Elliott KC et al. It's good to share: why environmental scientists' ethics are out of date. *Bioscience* 2015; 65: 69–73.
 - [14] Eschenfelder KR and Johnson A. Managing the data commons: controlled sharing of scholarly data. *J Assoc Inf Sci Tech* 2014; 65: 1757–1774.
 - [15] Candela L, Castelli D, Manghi P et al. Data journals: a survey. *J Assoc Inf Sci Tech* 2015; 66: 1747–1762.
 - [16] Nelson B. Data sharing: empty archives, <http://www.nature.com/news/2009/090909/full/461160a.html> (2009, accessed 08 June 2011)
 - [17] Kennan MA, Williamson K and Johanson G. Environmental voluntary groups: towards curating data for sharing, access and preservation. *P Am Soc Inform Sci Tech* 2013; 50: 1–10.
 - [18] Hannay T. From Web 2.0 to the global database. In: Hey T, Tansley S and Tolle K (eds) *The fourth paradigm: data-intensive scientific discovery*. Redmond, WA: Microsoft Research, 2009, pp. 215–220.
 - [19] Cragin MH, Palmer CL, Carlson JR et al. Data sharing, small science and institutional repositories. *Philos TR Soc A* 2010; 368: 4023–4038.
 - [20] Gardner D, Toga A, Ascoli G et al. Towards effective and rewarding data sharing. *Neuroinformatics* 2003; 1: 289–295.
 - [21] Piwowar HA and Chapman WW. A review of journal policies for sharing research data. In: *Open conference service, ELPUB*, Toronto, ON, Canada, 25–27 June 2008.
 - [22] Piwowar HA and Chapman WW. Public sharing of research datasets: a pilot study of associations. *J Informetr* 2010; 4: 148–156.
 - [23] Genova F, Arviset C, Almas BM et al. Building a disciplinary, world-wide data infrastructure. *Data Sci J* 2017; 16: 16.
 - [24] Kim Y and Burns CS. Norms of data sharing in biological sciences: the roles of metadata, data repository, and journal and funding requirements. *J Inf Sci* 2016; 42: 230–245.
-

-
- [25] Kim Y and Stanton JM. Institutional and individual factors affecting scientists' data-sharing behaviors: a multilevel analysis. *J Assoc Inf Sci Tech* 2016; 67: 776–799.
- [26] Tenopir C, Dalton ED, Allard S et al. Changes in data sharing and data reuse practices and perceptions among scientists world- wide. *PLoS ONE* 2015; 10: e0134826.
- [27] Shema H, Bar-Ilan J and Thelwall M. Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. *J Assoc Inf Sci Tech* 2014; 65: 1018–1027.
- [28] Kjellberg S. I am a blogging researcher: motivations for blogging in a scholarly context. *First Monday* 2010; 15, <https://first-monday.org/ojs/index.php/fm/article/view/2962/2580>
- [29] Letierce J, Passant A, Breslin J et al. Understanding how Twitter is used to spread scientific messages. In: *Proceedings of the WebSci10: extending the frontiers of society on-line*, Raleigh, NC, 26–27 April 2010, pp. 1–8, http://www.johnbreslin.org/files/publications/20100426_webs2010c.pdf
- [30] Eysenbach G. Can tweets predict citations? Metrics of social impact based on twitter and correlation with traditional metrics of scientific impact. *J Med Internet Res* 2011; 13: e123.
- [31] Priem J and Costello KL. How and why scholars cite on Twitter. *P Am Soc Inform Sci Tech* 2010; 47: 1–4.
- [32] Shema H, Bar-Ilan J and Thelwall M. Research blogs and the discussion of scholarly information. *PLoS ONE* 2012; 7: e35869.
- [33] Procter R, Williams R, Stewart J et al. Adoption and use of Web 2.0 in scholarly communications. *Philos T R Soc A* 2010; 368: 4039–4056.
- [34] Zhu Y. Who support open access publishing? Gender, discipline, seniority and other factors associated with academics' OA practice. *Scientometrics* 2017; 111: 557–579.
- [35] Braun V, Clarke V and Terry G. Thematic analysis. *Qual Res Clin Health Psychol* 2014; 24: 95–114.
- [36] Zeller R and Carmines E. *Statistical analysis of social data*. Chicago, IL: Rand McNally College Publishing Company, 1978.
- [37] Harrell FE. Binary logistic regression. In: Harrell FE (ed.) *Regression modeling strategies*. New York: Springer, 2015, pp. 219–274.
-