



This is a repository copy of *Publishers, funders and institutions: who is supporting UKRI-funded researchers to share data?*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/215258/>

Version: Published Version

Article:

Montague-Hellen, B. orcid.org/0000-0002-8159-9013 and Montague-Hellen, K. orcid.org/0000-0002-8159-9013 (2023) *Publishers, funders and institutions: who is supporting UKRI-funded researchers to share data?* *Insights*, 36. 4. ISSN 2048-7754

<https://doi.org/10.1629/uksg.602>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Publishers, funders and institutions: who is supporting UKRI-funded researchers to share data?

Researchers are increasingly being asked by funders, publishers and their institutions to share research data alongside written publications, and to include data availability statements to support their readers in finding this data. In the UK, UKRI (UK Research and Innovation) is one of the largest funding bodies and has had data-sharing policies for several years. This article investigates the reasons why a researcher may or may not share their data and assesses whether funders, publishers and institutions are supporting data-sharing behaviour through their policies and actions. A survey with 166 responses gave an indicative assessment of researcher opinions around data sharing, and a corpus of 3,277 journal articles retrieved from four UK institutions was analysed using multivariate logistic regression models to provide empirical evidence as to researcher behaviour around data sharing. The regression models provide insight into how this is affected by the funder, institution and publisher of the research. This study identifies that those publishers and funders who give clear guidance in their policies as to which data should be shared, and where this data should be shared, are most likely to encourage good practice in researchers.

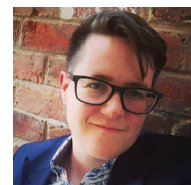
Keywords

open research, open data, data availability statements, UKRI, policy, publishers

Introduction

Open research, and the sharing of data, has rapidly grown as a practice during the twenty-first century.¹ Advocates laud the practice as one which speeds up the progress of research and ensures that public funds are used to the best effect.² As researchers have become increasingly concerned with ensuring that their data and findings are reproducible³ and that they follow responsible research practices,⁴ data sharing has become more popular. However, there are also detractors who highlight the extra costs and skills needed to openly share data, and who point to the risks in releasing data which could contain sensitive information, or which might be misinterpreted.⁵

Although there are many 'soft' reasons for researchers to share their data and to engage with open research practices, there have been few concrete incentives from funders, publishers or academic institutions.⁶ In 2011 the Research Councils UK (RCUK) Common Principles on Data Policy was released, a best practice guide to managing and sharing data, but not a mandate requiring that UK researchers shared their data.⁷ In 2015, this was updated to the 'Guidance on best practice in the management of research data' document, a set of seven principles laying out the best practice in research data management. The first of these principles is, 'publicly funded research data are a public good, produced in the public interest which should be made openly available with as few restrictions as possible in a timely and responsible manner'.⁸ This was a strongly worded encouragement to researchers funded by RCUK, later UKRI (UK Research and Innovation), that they should share their data. However, there was still a lack of concrete benefits or sanctions.



BETH MONTAGUE-HELLEN

Head of Library and Information Services
The Francis Crick Institute
(Previously: Senior Research Librarian
University of Nottingham)



KATE MONTAGUE-HELLEN

Research Sister
University of Sheffield

2 Many advocates believe that the compliance-based 'stick to punish the donkey' approach to open research is counterproductive and teaches researchers to simply do the minimum to complete a tick box.⁹ Instead of a compliance-based approach, an alternative way to encourage open research and data sharing is a reward-based approach, the 'carrot to encourage the donkey'. The Dutch Research Council set up a new fund in 2020 to promote open research, and this along with promotion incentives could help encourage open research, although these approaches are still relatively uncommon.¹⁰ In reality, UKRI uses neither the stick nor the carrot approach and leaves it up to the researcher's own ethical code to comply with policies. This approach is unlikely to lead to a high level of compliance.¹¹

'an alternative way to encourage open research and data sharing is a reward-based approach'

A data availability statement is one pillar of data-sharing practice and policy which has been promoted by funders, institutions and publishers. Data availability (or access) statements exist to tell the reader of an article how, or if, the underlying data can be accessed but do not force the researcher into robust data-sharing practices.¹² The most frequent content of data availability statements falls into three groups: data is in a repository, data is available on request, and the data can be found in the article or supplementary files.¹³ Guidance and support are given by funders, journals and institutions into the contents of these statements,¹⁴ but there appears to be little oversight into whether this guidance is being followed.

'UKRI uses neither the stick nor the carrot approach and leaves it up to the researcher's own ethical code'

Previous analysis of data availability statements and data-sharing behaviour has frequently concentrated on changes of behaviour elicited by policies enacted by a single journal e.g. PLOS ONE¹⁵ and the British Medical Journal,¹⁶ or changes seen in a selection of high impact journals.¹⁷ Recent research has shown that journals with strong data access policies have high compliance regarding the existence of a data availability statement, but that journal policies may be standardizing the use of statements which do not make it easy for other researchers to obtain the data.¹⁸ When approached for data, studies have discovered that researchers are often loath to provide data, regardless of the assertions made in their data availability statements.¹⁹

Longitudinal surveys such as the State of Open Science report,²⁰ and research conducted by DataOne,²¹ show a gradual increase in willingness to share data, with some differences based on age or discipline. In 2011, medicine was identified as a subject where authors were unlikely to share data.²² In the last few years however, many in the field of medicine, and other allied fields, have rapidly utilized the sharing of data to support the management of the Covid-19 pandemic.²³ It is important to note that although the pandemic may have increased the willingness of some researchers to share their data, it has not overcome all of the barriers that have previously prevented this, particularly in countries which lack the infrastructure or funds to support open research.²⁴

'studies have discovered that researchers are often loath to provide data'

What are the UKRI data requirements for data sharing?

UKRI is a governmental umbrella body which includes seven funding councils directly funding research in the UK: Arts and Humanities Research Council (AHRC), Biotechnology and Biological Sciences Research Council (BBSRC), Economic and Social Research Council (ESRC), Engineering and Physical Sciences Research Council (EPSRC), Medical Research Council (MRC), Natural Environment Research Council (NERC), Science and Technology Facilities Council (STFC). In the financial year 2019–2020, 3,829 grants and 441 fellowships were awarded. £3.28 billion was invested into UK research ideas and a further £1.72 billion in infrastructure.²⁵

'the pandemic may have increased the willingness of some researchers to share their data'

The overarching UKRI data policy is based on the 2015 'Guidance on best practice in the management of research data'.²⁶ Seven principles are outlined, and each of the individual

3 research councils has signed up to these principles, although in several cases they are built upon, or altered through, specific council guidance. The principles can be condensed into the following requirements:

- data should be shared as openly as possible
- data should be shared with as few restrictions as possible
- data should be shared in a timely manner
- data with long-term value should be preserved
- metadata should be included alongside data
- published results should always include a data availability statement
- a short time of privileged use is allowed.

Although data shared openly in a repository is considered to be the gold standard here, there are a number of reasons why researchers may feel that they cannot share their data. Very few of these reasons are considered acceptable according to UKRI policy, for example if the paper is a review article there is likely to be no data associated with it. UKRI policy allows for compliance where there is a need for sensitivity, or continued use by the research team, by the creation of metadata pages describing the data and how, or if, the data can be accessed. Therefore, a lack of associated data means the article is not considered compliant with UKRI policy.

For the purposes of this research, we have taken the following as a guide to whether researchers understand and comply with the spirit, in addition to the letter, of UKRI funding policies:

1. All published results contain a data availability statement.
2. 'Data available on request', or similar wording, has not been used.
3. Data is shared in a repository via a DOI, or where not available a direct link, accession number, or handle, is stated in the data availability statement.

To assess compliance, and the factors influencing and supporting researcher behaviour, a combination of survey data and empirical evidence from publications has been used. The survey is used as a general tool to gain an overall understanding of behaviours and beliefs around open research in UK researchers, a broad but shallow analysis. The empirical evidence from publications is then used as a deep but narrow analysis. The two analyses can be compared to give a greater insight to the researchers' understanding and behaviours.

Methods

Survey data

A survey of UK researchers was carried out between December 2021 and January 2022. The questions asked and the anonymized responses can be found in the accompanying open dataset.²⁷ Participants were invited through social media channels and through library communication channels. One hundred and twelve academic libraries in the UK were contacted and asked to share the survey with both their staff and student populations through whichever channels were deemed most appropriate.

The survey received 166 responses, a small number given the total number of researchers in the UK and the number of institutions contacted. Due to the number of responses, the results can only be considered indicative. Unfortunately, many of the institutions contacted did not manage to solicit responses from researchers. Often library communication channels are saturated and due to workload issues researchers may not respond to general calls.²⁸ An approach through academic department channels may have yielded more responses.

4 Researchers were asked who their research was funded by. Two datasets were created, one of UKRI-funded researchers (87), and a control dataset (79) where researchers were not funded by UKRI. Many of the UKRI researchers are funded by additional organizations, which may result in a lack of specificity in the results. Summary graphs were generated to show the distribution of responses to the questions:

- 'How well do you feel you understand your primary funder's policies on sharing data?'
- 'How appropriate and realistic do you think current data-sharing policies are for your field?'
- 'Have you deposited in (a) a subject repository, (b) an institutional repository (c) a general repository?'
- 'If you have openly shared data, what were your reasons?'
- 'If you have come across difficulties sharing your data, what were these?'

The percentage of survey respondents stating that they had (1) written a data availability statement, (2) used 'data available on request', (3) included a link to data, was used as a comparison to assess community reporting of behaviour compared to the observed journal article dataset. Further questions were asked in the survey, but they have not been analysed here.

To bolster the survey data created for this analysis further survey data has been obtained through the State of Open Data survey 2021.²⁹ This survey is part of a longitudinal study into open research behaviours: the survey has run for six years accruing 21,000 responses from 192 countries. The 2021 survey data contained a total of 4,491 responses. The data was filtered to only include responses from UK researchers, resulting in 197 responses. The data cannot be assumed to be independent of the data collected in this study's survey, and so the results have not been combined, rather reported alongside where it was possible to do so. Data on the funding of respondents has not been made publicly available, and so the dataset can only be used as an indication of the overall picture of UK research. Although the questions in the State of Open Data survey were not the same as those in the survey developed for this analysis, where the questions align, they have been included as a comparator to increase confidence in the data and conclusions. The questions regarding reasons for not sharing data (Q3.10), for depositing in a repository (Q3.4) and for whether they would like additional guidance on complying with policies (Q3.13) have been included.

Journal article dataset

A corpus of journal articles and conference proceedings with a publication date between January 2021 and February 2022 inclusive were collected from four Russell Group (RG) universities via their institutional repositories. The four Russell Group universities were mid-sized universities, ranked in the top 100 of the QS World University Rankings 2021 table,³⁰ all four conduct research across science, technology, engineering and mathematics (STEM) and arts, humanities and social sciences (AHSS) subjects, submitted similar numbers of researchers to the UK Research Excellence Framework (REF) and submitted to a wide range of Units of Assessment in REF2021.³¹ The universities were chosen to act as replicates, giving greater confidence to the analysis. Review articles and opinion pieces were not included in the dataset as these are unlikely to contain underlying data. Evidence has shown that the proportion of journal articles deposited in UK institutional repositories has rapidly increased year on year since the REF Open Access requirements were introduced in 2014.³² Ten Holter describes researchers as frustrated with the process of depositing in an institutional repository, but reports that repository managers believe they are predominantly compliant with the REF requirements.³³ Therefore, we believe that the period immediately prior to the submission of REF2021 is likely to provide a corpus containing the majority of the articles published during this time.

'the proportion of journal articles deposited in UK institutional repositories has rapidly increased'

5 The universities and the articles have been anonymized to prevent any negative consequences for institutions, or individual researchers, as a result of non-compliance with UKRI policy. All journal articles metadata-tagged as UKRI funded and 300 metadata-tagged as non-UKRI-funded articles were retrieved for each institution. Narrative reviews, opinion pieces and other types of article unlikely to contain underlying data were removed. The control sample was selected by setting the advanced search to select only items between the dates of interest in chronological order, removing those which were not research articles, or which were funded by UKRI. The articles were not matched for subject as this metadata was not included as standard in the repositories. However, a post hoc analysis shows that where a publisher occurs more than ten times in the overall dataset, it is always found in both the non-UKRI dataset and in the UKRI dataset. Informa UK (Taylor and Francis) is slightly overrepresented in the non-UKRI dataset. All institutions analysed had a transformative agreement deal with this publisher providing a no article processing charge (APC) route to publishing gold open access. This may be a factor, but it has not been seen in other publishers with transformative agreements so cannot be confirmed.

The resulting dataset contained 3,277 entries.³⁴ Some duplication was present in the dataset as the same article could be identified from more than one institution or funded by more than one UKRI research council, but 95% of the journal articles were unique. To remove the impact of double counting in the dataset, two additional derivative datasets were created: one which was non-redundant regarding funding councils, one which was non-redundant regarding institution.

Each article was manually assessed to determine whether there was a statement addressing the location of, or access to, the data. Statements which were not formal data availability statements were included provided they clearly described if the data was shared or where it could be found. Statements were identified by looking at data on the publisher site, in the abstract and at the start of the article, and in the end sections of the article, including the conclusions section. It is possible that further articles may have included statements within the methods section, but these were not included. Articles were searched on two independent occasions to verify the data.

'Articles were searched on two independent occasions to verify the data'

Data availability statements were checked in two stages, firstly for the occurrence of a statement and a direct link to a dataset or metadata record. If a statement was present, but there was no direct link, the statement was analysed for the reasons why this was missing and was coded as one or more of: data available on request, no data, sensitive data, general location but no link, technical concerns (large data, not enough time, ongoing research).

Data analysis

Description statistics were generated for the journal article dataset. A percentage value of compliance with (1) a data availability statement being present in an article, (2) the statement having a link to a dataset or metadata record, (3) the statement containing a 'data available on request' statement was calculated for each research council and for each of the four institutions. The mean and standard deviation of these results by research council were graphed and compared to the survey reported behaviour percentage values.

Multivariate logistic regression models were used to estimate the likelihood of (1) a data availability statement being present in an article, (2) the statement having a link to a dataset or metadata record, (3) the statement containing a 'data available on request' statement. Ordinal independent variables included institution retrieved from the journal publisher and funder. All independent variables were reordered to their modal value. Articles missing data availability statements were omitted from models 2 and 3. Models were run using the non-redundant datasets.

All models exhibited low multicollinearity. In model 1 autocorrelation was found amongst the independent variables, this may be due to the strong influence of field over both funder and publisher.

Results and discussion

Are funder data-sharing policies clear to researchers?

The small number of respondents to the survey means that the results may be biased, particularly towards those who have strong opinions either for or against open research. However, we believe that these results can still show indicative patterns of opinion. A majority of respondents to the survey described themselves as open research enthusiasts, with similar numbers across those funded by UKRI and those who were not funded by UKRI. It is unsurprising given the time and workload pressures on researchers that those who found the time to respond were particularly interested in the topic. The respondents were primarily from research-intensive universities and STEM fields where there may be a greater emphasis on open research and a greater understanding of the benefits and requirements. These demographics may mean that the results are more positive than those that would be found in a larger, more generalized population of researchers.

When asked about their understanding of their funder data-sharing policies, scored on a scale between 1 (low) and 10 (high) (Figure 1), UKRI-funded researchers within the subset of respondents to our survey scored their own understanding as higher than did those not funded by UKRI. The mean reported level of understanding for UKRI researchers was 8.4, the mean value for researchers not funded by UKRI was at 6.1. Overall, 33% of researchers rated their understanding as five or less. The State of Open Data survey did not ask researchers to rate their understanding, but it did ask if researchers would like more guidance. Thirty-nine per cent of researchers stated that they would like a greater level of guidance on how to comply with funder policies, a similar level to the proportion of those who rated their understanding as less than five. Ratings of the appropriateness of policies (see Figure 1) between UKRI and non-UKRI researchers showed no difference between the two groups, with the overall mean at 6.2.

'The respondents were primarily from research-intensive universities and STEM fields'

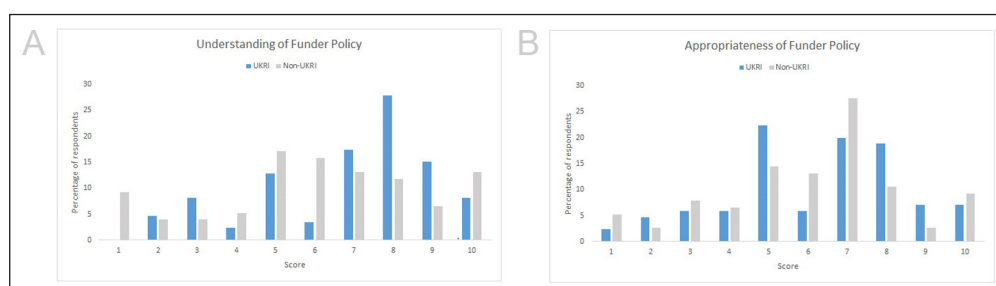


Figure 1. Two bar charts showing researcher's self-reported understanding of funder data-sharing policies (scale 1 – 10) and their rating of their appropriateness for the researcher's field (scale 1 – 10). Respondents have been split by whether they are funded by UKRI or not

Something that should be considered regarding this data is the seniority of the researchers. The UKRI-funded respondents were more likely to have been a researcher for over ten years, possibly due to the difficulties for junior researchers in obtaining independent funding and the likelihood of senior researchers having grants from a range of publishers. Older researchers in the sample had a higher level of understanding of policies than younger researchers, however the difference between groups was not as great as when comparing the difference between the UKRI- and non-UKRI-funded groups.

The survey asked whether researchers had previously submitted data to a repository. Overall, only 65% of respondents had deposited data in a repository (subject, institutional, or general repositories). The State of Open Data survey reported a slightly greater proportion at 71% (with slightly different types of repositories reported: institutional, funder or general). Although the data shows that the majority of researchers have deposited data in a

'Older researchers in the sample had a higher level of understanding of policies'

7 repository, there is still a sizeable number of researchers who are yet to do so. For all three types of repositories, UKRI researchers were more likely to have deposited their data than those who were not funded by UKRI (see Figure 2). This was particularly the case for institutional repositories. A further 20% of responders in our survey reported that they would like to deposit but so far have not. It was unclear whether these researchers have had the opportunity to deposit previously, or whether they were still to publish for the first time. Given the right support, it is likely that many of the researchers in this category could become data-sharing advocates and practitioners in the future.

'the majority of researchers have deposited data in a repository'

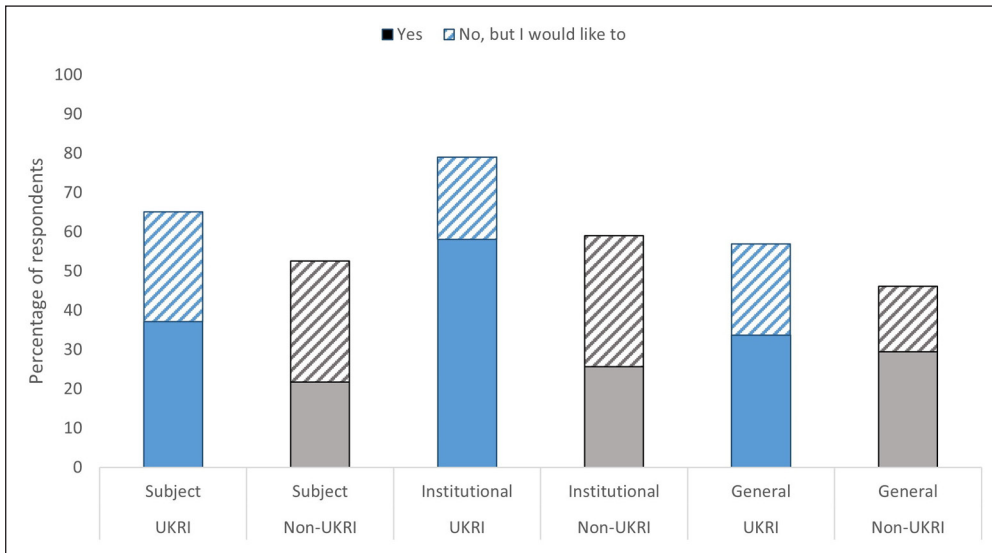


Figure 2. Bar chart showing the percentage of researchers self-reporting as having deposited in an institutional repository. Respondents have been split by whether they are funded by UKRI or not

The survey also asked respondents to give their reasons for sharing data, allowing each respondent to give multiple reasons (See Figure 3). Although low (<50%), the number of researchers who gave 'required by funder' as a reason for sharing data was higher for the UKRI-funded researchers than for other researchers. It is unclear whether this was a lack of knowledge on the part of the respondents or whether they simply did not consider it to be a high priority reason. Funder requirements was provided as a reason for sharing less often than sharing with the public or increasing the impact of research. This may show that for some researchers, although they are required to share data by their funder, this is not a primary driver for them. However, given the small sample size, further research would be needed to confirm this.

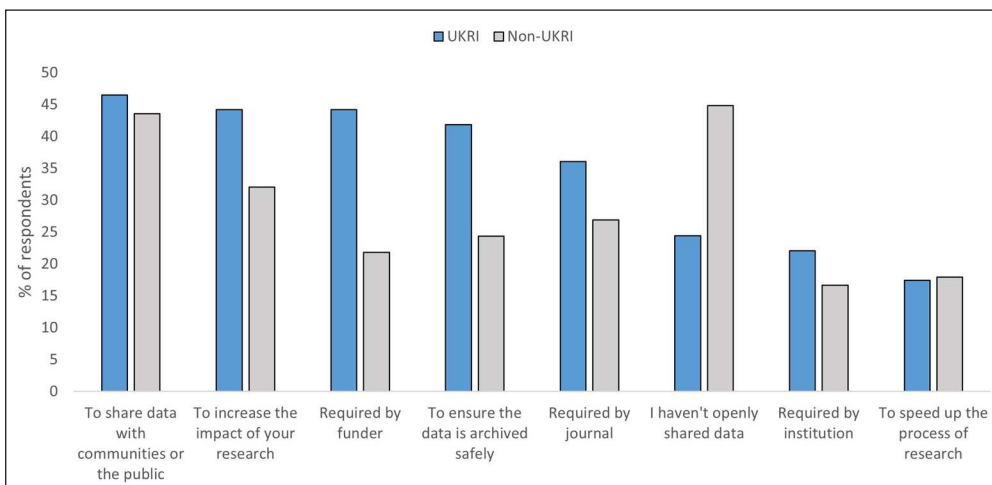


Figure 3. Bar chart showing the percentage of researchers giving different reasons for sharing their data. Respondents have been split by whether they are funded by UKRI or not

8 In addition to asking why researchers shared their data, the survey also asked why a researcher would not share, or may find it difficult to share, data (See Figure 4). This question, while not the same as that used by the State of Open Data survey, does allow for some benchmarking across surveys. In that dataset, concerns about sensitive or personal data were not considered to be such a great barrier and neither was lack of a suitable repository. Costs, however, were a higher concern. Lack of time and lack of support were the most frequently given reasons for UKRI-funded researchers, suggesting that any non-compliance was not due to a lack of understanding of policies, or a concern about the data, but an inability to manage this additional workload.

'non-compliance was not due to a lack of understanding of policies ... but an inability to manage this additional workload'

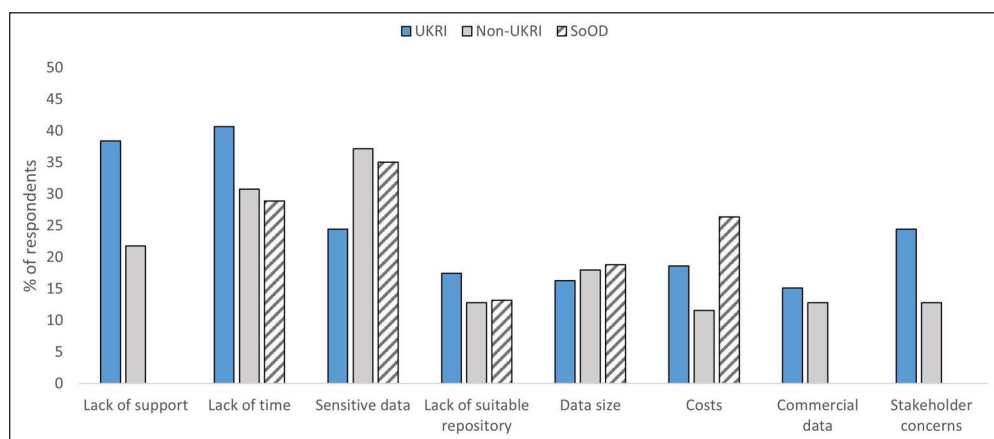


Figure 4. Percentage of researchers giving each reason why they would not share their data. Respondents have been split by whether they are funded by UKRI or not. The data from the State of Open Data survey has been included where questions aligned – there was no equivalent question about lack of support, commercial data or stakeholder concerns

The responses to the survey consistently showed a more positive relationship to open research and data sharing from the UKRI-funded researchers than it did from the non-UKRI-funded researchers. However, due to the small number of respondents, this data may be biased and may not accurately describe the opinions and beliefs of researchers more widely. The second part of this study attempts to empirically assess the behaviours of UKRI- and non-UKRI-funded researchers and to determine where pressure and support for open research practices comes from.

Does self-perception of compliance and understanding match reality?

Both the survey created for this study and the State of Open Data suffer from a lack of responses to enable generalizations to be made. However, for certain UKRI requirements, we can use empirical data to map behaviour. Here we have analysed journal articles from 2021 and early 2022, identified through the institutional OA repositories of four Russell Group universities. Each article was manually assessed to identify any statements pertaining to the location or access of data.

Criterion 1: all published results contain a data availability statement

Although some research councils, particularly EPSRC and BBSRC, allow for author judgement about whether a data availability statement needs to appear in a publication, the overarching requirement from the Common Principles on Research Data is that 'Published results should always include information about how to access the supporting data'.³⁴

A higher percentage of researchers who were funded by the UKRI self-reported as having included a data availability statement in previously published articles (45% UKRI, 38% non-UKRI). Although the survey sample size is small, and therefore may be biased, these reported percentages mapped closely to the overall percentage of articles with data

9 availability statements in the repository corpora, although considerable variability between different funding councils can be seen (see Figure 5). Only those articles funded by AHRC were less likely to contain a data availability statement than those which were not funded by UKRI.

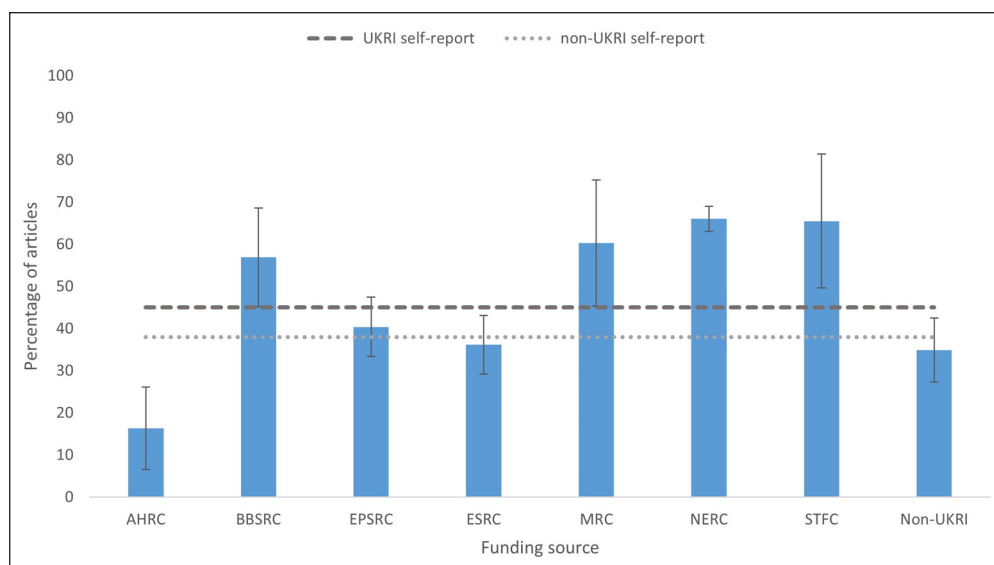


Figure 5. Percentage of articles in dataset which contain data availability statements, using the non-redundant funder dataset. Data from each institution was calculated separately and error bars represent one standard deviation from the mean of these results. Dashed lines show the percentage of respondents in the survey self-reporting as having written a data availability statement

Multivariate logistic regression models (see Table 1) show that articles funded by BBSRC, MRC, NERC, EPSRC and STFC were all statistically more likely to contain a data availability statement than an article which was not funded by UKRI. Articles funded by NERC were the most likely to contain a statement which may be due to the extra support and direction NERC researchers are provided with when depositing into the NERC funded data repositories.

Fifteen of the publisher groups were identified as being statistically more or less likely to have a data availability statement than the reference group of Elsevier, chosen due to the large number of articles and the lack of a strong data availability statement policy. Although articles published with 14 of the publishers were more likely to contain a data availability statement, Institute of Electrical and Electronics Engineers (IEEE) articles were significantly less likely to.

Of the 14 journals which showed a significantly improved likelihood of containing a data availability statement, eight of these mandated statements within all publications (Frontiers,³⁵ PLoS,³⁶ BMJ,³⁷ MDPI,³⁸ Nature,³⁹ American Society for Microbiology,⁴⁰ Optical Society for America,⁴¹ IOP⁴²), including specific sections of the article for these statements and often including it within the submission workflow. A further four publishers (OUP,⁴³ Springer,⁴⁴ EDP Sciences,⁴⁵ Wiley Blackwell⁴⁶) strongly encouraged data availability statements and provided guidance but stopped short of mandating these for all journals.

The IEEE does not mention data availability statements within their submission guidance for authors, they provide no encouragement to authors to include one and force authors to find a location in the article to include it as no specific section is provided. It is therefore not surprising that IEEE articles were less likely to contain a statement than the Elsevier reference group, as Elsevier does at least provide information about data availability statements, even if they are not frequently mandated in their journals. We were unable to find any guidance provided for a further two publishers (the American Astronomical Society

10 and the Institute of Mathematical Statistics) about including a data availability statement, although their articles showed an increased likelihood of providing these statements – implying that the guidance found within instructions to authors is not the only factor at play. Instead, it may be that field or funder related practices and policies have encouraged the inclusion of data availability statements in these articles.

A significant difference was found between institutions RG2 and RG4, with articles from RG2 being more likely to contain a data availability statement, but not between any other pair of institutions. This may signify that individual institutional policies, and the quality of support at the respective institutions, can have an effect on the occurrence of a data availability statement. However, the institutions chosen were similar, all research-intensive institutions from the Russell Group, and are likely to be providing an equivalent level of support. Without further research it is not possible to identify the reasons for any difference in compliance between institutions.

	Comparison	Direction	Exp odds	P value
<i>Funding source</i>				
NERC	Non-UKRI	+	2.57	1.06e-5
MRC	Non-UKRI	+	2.27	8.29e-6
BBSRC	Non-UKRI	+	2.05	4.32e-5
STFC	Non-UKRI	+	1.93	6.09e-3
EPSRC	Non-UKRI	+	1.45	3.42e-3
<i>Publisher</i>				
Frontiers Media SA	Elsevier	+	383	4.32e-9
PLoS	Elsevier	+	190	2.51e-7
BMJ Publishing Group	Elsevier	+	21.5	1.51e-16
MDPI AG	Elsevier	+	20.7	9.97e-48
Nature Publishing Group	Elsevier	+	13.8	2.62e-37
American Astronomical Society	Elsevier	+	7.88	1.42e-6
Oxford University Press	Elsevier	+	7.78	1.37e-20
IOP	Elsevier	+	6.5	4.59e-9
American Society for Microbiology	Elsevier	+	5.49	2.43e-2
Springer	Elsevier	+	5.2	7.09e-22
Optical Society of America	Elsevier	+	5.0	3.64e-2
Institute of Mathematical Statistics	Elsevier	+	4.79	4.76e-2
EDP Sciences	Elsevier	+	4.66	1.58e-3
Wiley Blackwell	Elsevier	+	3.77	7.02e-17
IEEE	Elsevier	-	0.45	2.25e-2
<i>Institutions</i>				
RG2	RG4	+	1.32	4.70e-2
RG4	RG2	-	0.76	4.70e-2

Table 1. Multivariate logistic regression analysis results assessing the likelihood of an article containing a data availability statement based on (a) the funder, (b) the publisher and (c) the institution

Criterion 2: 'data available on request' has not been used

The second criterion which covers most of the UKRI funding councils is that 'data available on request' statements are not considered sufficient. Having to request data cannot be seen as being as open as possible or as having as few restrictions as possible. As a minimum, a link should be provided to a metadata page, even if the data itself cannot be obtained without contacting the authors or institutions. The funding council that deviates the most from this is the BBSRC,

“data available on request” statements are not considered sufficient’

11 which specifically allows for data to be provided via a direct request to the author.⁴⁷

By contrast, NERC state that a DOI to their recommended repository should be included when used.⁴⁸

Non-UKRI authors showed the highest frequency of using data available on request statements, and additionally showed a higher self-reporting level for this method of data sharing (see Figure 6). Although BBSRC specifically allows for data to be available only via contacting authors, this kind of statement still occurs less frequently than for non-UKRI authors.

'Non-UKRI authors showed the highest frequency of using data available on request statements'

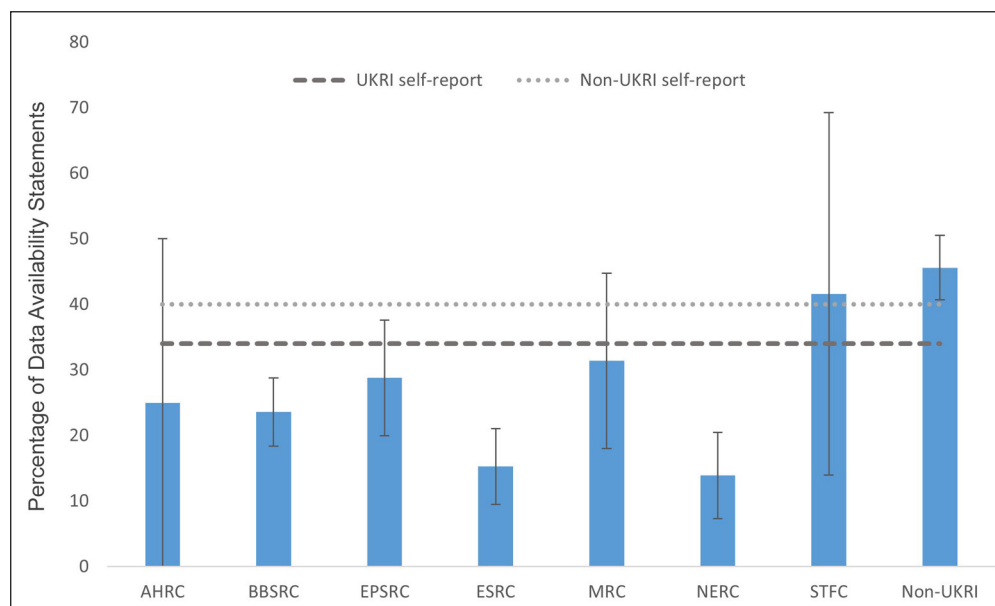


Figure 6. Percentage of articles in the dataset which contain data availability statements with 'available on request', 'contact authors' or similar using the non-redundant funder dataset. Data from each institution was calculated separately and error bars represent one standard deviation from the mean of these results. Dashed lines show the percentage of respondents in the survey self-reporting as having previously included an 'available on request' statement in an article

More variation was seen between the institutions when analysing the likelihood of a data available on request statement occurring (see Table 2). A significant difference was seen between RG1 and RG2, and RG2 and RG4, with RG2 showing the lowest likelihood. There may be confounding factors such as the disciplinary profile of the institutions and choices to where articles are published, however, these results imply that institutional guidance in policy and advice discourages researchers from relying on the data on request option within some universities.

Articles funded by four of the research councils were less likely to use available on request statements than articles that were not funded by UKRI: BBSRC, EPSRC, ESRC and NERC. Despite BBSRC funded authors being explicitly allowed, even recommended, to use author contact as a means of sharing data, these articles were still significantly less likely to contain this type of statement than the non-UKRI-funded articles.

Only five publishers showed a significant difference from the Elsevier reference (see Table 2). Frontiers, IOP, BMJ and Springer were all more likely to contain the data available on request statement than Elsevier. Only articles published by The Royal Society were less likely to contain this type of statement and this was a very small effect size. Although many journals appear to be successful in encouraging the occurrence of a data availability statement, it does not appear that the researchers are encouraged to make retrieving the underlying data easier for their readers.

Frontiers,⁴⁹ IOP⁵⁰ and BMJ⁵¹ are three of the publishers which showed an increased likelihood of containing a 'data available on request statement'. All three require a data availability statement as part of their submission process, forcing the authors to engage with the process. However, one of the recommended wordings is 'data available on request',

12 allowing authors to comply with the statement requirement but without encouraging engagement with open research practices. The Royal Society by contrast does not allow for a simple data available on request statement and instead requires that data is deposited in a repository and publicly available. It appears that this policy is actively supporting authors in choosing not to use a 'data available on request' statement.

	Comparison	Direction	Exp odds	P value
<i>Funding source</i>				
EPSRC	Non-UKRI	-	0.58	2.84e-3
BBSRC	Non-UKRI	-	0.30	4.3e-5
NERC	Non-UKRI	-	0.29	1.61e-5
ESRC	Non-UKRI	-	0.21	8.10e-5
<i>Publisher</i>				
Frontiers Media SA	Elsevier	+	7.6	1.68e-10
IOP	Elsevier	+	3.59	3.01e-3
BMJ Publishing Group	Elsevier	+	2.51	1.55e-2
Springer	Elsevier	+	1.77	3.88e-2
Royal Society	Elsevier	-	0.08	1.49e-2
<i>Institutions</i>				
RG2	RG1	-	0.60	5.57e-3
RG1	RG2	+	1.67	5.57e-3
RG4	RG2	+	1.51	3.5e-2
RG2	RG4	-	0.66	3.49e-2

Table 2. Multivariate logistic regression analysis results assessing the likelihood of a data availability statement within an article containing 'data available on request' based on (a) the funder, (b) the publisher and (c) the institution.

Criterion 3: data is shared in a repository

NERC funded authors are the most likely to have deposited in a repository (see Figure 7). One explanation for this may be that NERC gives clear instructions and support for depositing certain types of data into specified data repositories⁵² and this may be leading to a higher level of compliance. This criterion shows the greatest difference between the self-reported values and the percentages seen in the 2021-2022 article corpora. However, the survey question did not ask if the repository had been reported in a data availability statement, just that data had been deposited.

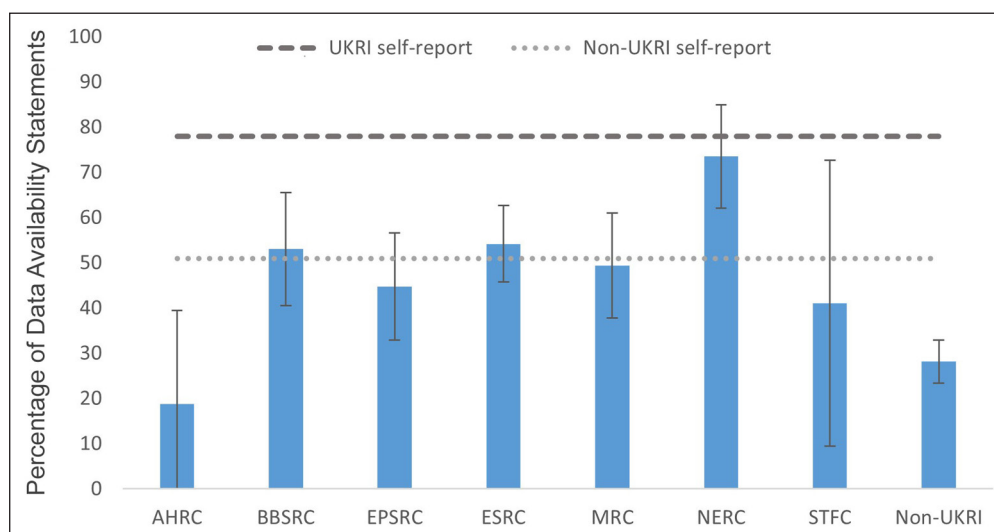


Figure 7. Percentage of articles in the non-redundant funder dataset which contain data availability statements with direct links to datasets or metadata records. Data from each institution was calculated separately and error bars represent one standard deviation from the mean of these results. Dashed lines show the percentage of respondents in the survey self-reporting as having previously deposited in a repository.

13 The regression analysis (Table 3) shows that articles funded by four of the research councils were more likely to contain a direct link to the research data or to a metadata page describing the data, than those not funded by UKRI. These councils were the same as those whose articles were more likely to contain a data availability statement: BBSRC, EPSRC, ESRC and NERC. NERC funded articles were over four times more likely than the non-UKRI articles to have a direct link to data, and ESRC articles were over three times as likely. One of the major differences between these councils and the other UKRI councils is that NERC and ESRC data policies specify, or recommend, repositories for the submission of data. NERC also, crucially, provides support during the deposit process. This extra support and guidance may be helping to foster a change in the behaviour of researchers around data deposit and data availability statements.

'The regression analysis shows that articles funded by four of the research councils were more likely to contain a direct link to the research data'

Articles from RG1 were significantly less likely to contain a link to a repository than those from RG2 and RG3. This implies that the structures and guidance in place at some institutions may be more successful at supporting researchers than at others.

Four of the publishers were significantly less likely to have articles with direct links: MDPI, Frontiers, BMJ Publishing Group and IOP. Five of the publishers were more likely to have direct links within their data availability statement: American Chemical Society, AAAS, The Royal Society of Chemistry, EDP Sciences and Nature Publishing Group. The publishers who were most likely to have direct links were all STEM publishers, with two from Chemistry, and this result may be partially related to norms within the fields. However direct publisher policies and guidance may also be creating this effect, for example, AAAS,⁵³ Royal Society of Chemistry⁵⁴ and Nature⁵⁵ publishers all provide a clear list of recommended or mandated repositories for researchers to use. While the American Chemical Society⁵⁶ does not provide this level of guidance, it does point authors towards resources to help them source a repository. The publishers where links to repositories were less likely to occur mostly have far less detailed guidance on repositories for authors to use. The exception to this was Frontiers which does mandate repositories for certain data,⁵⁷ however, due to the large number of topics that Frontiers publishes, these repositories only cover a fraction of the data associated with their articles.

	Comparison	Direction	Exp odds	P value
<i>Funding source</i>				
NERC	Non-UKRI	+	4.23	1.75e-8
ESRC	Non-UKRI	+	3.45	8.73e-5
BBSRC	Non-UKRI	+	1.90	4.07e-3
EPSRC	Non-UKRI	+	1.83	1.04e-3
<i>Publisher</i>				
MDPI	Elsevier	-	0.39	5.04e-4
Frontiers Media SA	Elsevier	-	0.38	2.75e-3
IOP	Elsevier	-	0.27	9.32e-3
BMJ Publishing Group	Elsevier	-	0.26	9.95e-3
American Chemical Society	Elsevier	+	16.53	2.31e-4
AAAS	Elsevier	+	14.43	1.35e-2
Royal Society of Chemistry	Elsevier	+	8.45	8.08e-3
EDP Sciences	Elsevier	+	6.52	7.55e-3
Nature Publishing Group	Elsevier	+	2.07	3.98e-3
<i>Institutions</i>				
RG2	RG1	+	1.56	1.72e-2
RG3	RG1	+	1.55	2.51e-2
RG1	RG2	-	0.64	1.72e-2
RG1	RG3	-	0.65	2.51e-2

Table 3. Multivariate logistic regression analysis results assessing the likelihood of a data availability statement within a link to an associated data deposit, or metadata page, based on (a) the funder, (b) the publisher and (c) the institution

Conclusions

UKRI funding correlates positively with the occurrence and quality of data availability statements within journal articles, at least for the STEM-based councils and within the constraints of our dataset. However, even in the STEM disciplines it is clear that the existence of a policy requiring the presence of a data availability statement is not enough to elicit a behaviour change in all researchers. Without rewards or sanctions these policies still rely on the goodwill, enthusiasm, time and resources of the researcher. UKRI-funded researchers were most likely to cite lack of time or lack of support as their reasons for not sharing, and without making this process easier for researchers, it may be difficult to elicit further change.

'UKRI funding correlates positively with the occurrence and quality of data availability statements'

The inclusion of data availability statements has still not been normalized in AHSS research to the extent seen in STEM fields. This can be seen in AHRC and ESRC funded research having similar or lower frequencies of data availability statements to those from researchers across all fields who were not funded by UKRI. In these cases, the funding policies appear to have had little impact. The ESRC guides researchers towards the UK Data Archive as the best place for depositing social sciences data, and where a data availability statement exists this does appear to have encouraged direct links to data. However, ESRC researchers are still much less likely to include a data availability statement in the first place.

'The inclusion of data availability statements has still not been normalized in AHSS research'

Journals and publishers clearly have a strong influence over data availability statement behaviours. Where journals make the statement compulsory, the occurrence increases, but in many cases the statement still fails to direct the reader to the data. Where journals make it more difficult to include a statement, through the lack of guidance and encouragement, they can reduce the likelihood of any data availability statement appearing in the article at all. However, where journal mandates for data deposit exist, with clear guidance and direction, publisher policies can be clearly seen to have a strong effect on behaviour. Good examples of these types of publishers are AAAS and the Royal Society of Chemistry, which provide guidance, recommended repositories and a requirement to deposit.

'Journals and publishers clearly have a strong influence over data availability statement behaviours'

Although there are signs that funder policies and mandates can influence behaviour, these policies alone do not appear to have created a strong change. Field norms and journal policies still clearly influence researcher behaviour in this area, as may support from institutions. The new UKRI open access policy contains a stronger stipulation for data availability statements, and it may be hoped that this will support a change in behaviour in AHSS subjects and continue to support STEM subjects in sharing data. However, current evidence around policies which require a data availability statement but provide no further guidance, shows that this is unlikely and may instead be seen as a box-ticking exercise. The strength of adherence to policies by NERC funded authors, and those publishing with certain publishers such as AAAS, instead provides an insight into how to elicit change. This change requires clear guidance, explicit instructions and associated support provided for researchers.

'Field norms and journal policies still clearly influence researcher behaviour'

CRedit

Beth Montague-Hellen: conceptualization, methodology, formal analysis, writing – original draft, writing – review and editing

Kate Montague-Hellen: methodology, formal analysis, writing – review and editing

Data accessibility statement

Anonymized survey data is available under a CC BY licence at <http://doi.org/10.17639/nott.7214>. Anonymized article corpus data is available under a CC BY licence at <http://doi.org/10.17639/nott.7214>. The State of Open Data survey results can be retrieved from <https://doi.org/10.6084/m9.figshare.17061347.v1>

Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other *Insights* articles can be accessed here – click on the URL below and then select the 'full list of industry A&As' link: <http://www.uksg.org/publications#aa>.

Competing interests

The authors have declared no competing interests.

References

1. "The State of Open Data 2021," Digital Science et al., DOI: <https://doi.org/10.6084/m9.figshare.17061347.v1> (accessed 4 January 2023).
2. Christopher Allen and David M. A. Mehler, "Open science challenges, benefits and tips in early career and beyond," *PLOS Biology* 17, no. 5 (2019): e3000246, DOI: <https://doi.org/10.1371/journal.pbio.3000246> (accessed 4 January 2023).
3. Marcus R. Munafò et al., "A manifesto for reproducible science," *Nature Human Behaviour* 1, no. 1 (2017): 0021, DOI: <https://doi.org/10.1038/s41562-016-0021> (accessed 4 January 2023).
4. Joeri K. Tijndik et al., "Towards a research agenda for promoting responsible research practices," *Journal of Empirical Research on Human Research Ethics* 16, no. 4 (2021): 450–460, DOI: <https://doi.org/10.1177/15562646211018916> (accessed 4 January 2023).
5. Marijn Janssen, Yannis Charalabidis, and Anneke Zuiderwijk, "Benefits, adoption barriers and myths of open data and open government," *Information systems management* 29, no. 4 (2012): 258–268, DOI: <https://doi.org/10.1080/10580530.2012.716740> (accessed 4 January 2023).
6. Brian A. Nosek et al., "Promoting an open research culture," *Science* 348, no. 6242 (2015): 1422–1425, DOI: <https://doi.org/10.1126/science.aab2374> (accessed 4 January 2023).
7. Sarah Jones, "Developments in research funder data policy," *International Journal of Digital Curation* 7, no 1 (2012), DOI: <https://doi.org/10.2218/ijdc.v7i1.219> (accessed 4 January 2023).
8. "Guidance on best practice in the management of research data," United Kingdom Research Councils, 2015, <https://www.ukri.org/publications/guidance-on-best-practice-in-the-management-of-research-data/> (accessed 12 January 2023).
9. Carolyn Ten Holter, "The repository, the researcher, and the REF: 'It's just compliance, compliance, compliance,'" *The Journal of Academic Librarianship* 46, no. 1 (2020): 102079, DOI: <https://doi.org/10.1016/j.acalib.2019.102079> (accessed 4 January 2023).
10. Hans de Jonge, Maria Cruz, and Stephanie Holst, "Funders need to credit open science," *Nature* 599, no. 7885 (2021): 372–372, DOI: <https://doi.org/10.1038/d41586-021-03418-1> (accessed 4 January 2023).
11. R. Kent Weaver, "Compliance regimes and barriers to behavioral change," *Governance* 27, no. 2 (2014): 243–265, DOI: <https://doi.org/10.1111/gove.12032> (accessed 4 January 2023).
12. Giovanni Colavizza et al., "The citation advantage of linking publications to research data," *PLoS ONE* 15, no. 4 (2020): e0230416, DOI: <https://doi.org/10.1371/journal.pone.0230416> (accessed 4 January 2023).
13. Colavizza et al., "The citation advantage," e0230416.
14. Rebecca Grant and Iain Hrynaszkiwicz, "The impact on authors and editors of introducing Data Availability Statements at Nature journals," *International Journal of Digital Curation* 13, no.1 (2018), DOI: <https://doi.org/10.2218/ijdc.v13i1.614> (accessed 4 January 2023).
15. Lisa M Federer et al., "Data sharing in PLOS ONE: an analysis of data availability statements," *PLoS ONE* 13, no. 5 (2018): e0194768, DOI: <https://doi.org/10.1371/journal.pone.0194768> (accessed 4 January 2023).
16. Anisa Rowhani-Farid and Adrian G. Barnett, "Has open data arrived at the *British Medical Journal (BMJ)*? An observational study," *BMJ Open* 6, no. 10 (2016): e011784, DOI: <https://doi.org/10.1136/bmjopen-2016-011784> (accessed 4 January 2023).
17. Alawi A Alsheikh-Ali et al., "Public availability of published research data in high-impact journals," *PLoS ONE* 6, no. 9 (2011): e24357, DOI: <https://doi.org/10.1371/journal.pone.0024357> (accessed 4 January 2023).
18. Colavizza et al., "The citation advantage," e0230416.
19. Mirko Gabelica, Ružica Bojčić, and Livia Puljak, "Many researchers were not compliant with their published data sharing statement: mixed-methods study," *Journal of Clinical Epidemiology* (2022), DOI: <https://doi.org/10.1016/j.jclinepi.2022.05.019> (accessed 4 January 2023).
20. Digital Science et al., "The State of Open Data 2021."
21. Carol Tenopir et al., "Changes in data sharing and data reuse practices and perceptions among scientists worldwide," *PLoS ONE* 10, no. 8 (2015): e0134826, DOI: <https://doi.org/10.1371/journal.pone.0134826> (accessed 4 January 2023).
22. Carol Tenopir et al., "Data sharing by scientists: practices and perceptions," *PLoS ONE* 6, no. 6 (2011): e21101, DOI: <https://doi.org/10.1371/journal.pone.0021101> (accessed 4 January 2023).

23. Peter W Harrison et al., "The COVID-19 Data Portal: accelerating SARS-CoV-2 and COVID-19 research through rapid open access data sharing," *Nucleic Acids Research* 49, no. W1 (2021): W619–W623, DOI: <https://doi.org/10.1093/nar/gkab417> (accessed 4 January 2023).
24. Sarah Wehbe, et al., "COVID-19 in the Middle East and North Africa region: an urgent call for reliable, disaggregated and openly shared data," *BMJ Global Health* 6, no. 2 (2021): e005175, DOI: <https://doi.org/10.1136/bmjgh-2021-005175> (accessed 4 January 2023).
25. "UKRI Annual Report and Accounts 2019–2020," UKRI, <https://www.ukri.org/wp-content/uploads/2020/10/UKRI-050920-AnnualReport2019-2020.pdf> (accessed 4 January 2023).
26. "Guidance on best practice in the management of research data," UKRI, 2015, <https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-GuidanceBestPracticeManagementResearchData.pdf> 08/04/22 (accessed 4 January 2023).
27. Beth Montague-Hellen and Kate Montague-Hellen, "Researcher attitudes to data sharing 2022," 2022, DOI: <http://doi.org/10.17639/nott.7214> (accessed 12 January 2023).
28. Barry Bozeman and Jan Youtie, "Robotic bureaucracy: Administrative burden and red tape in university research," *Public Administration Review* 80, no. 1 (2020), 157–162, DOI: <https://doi.org/10.1111/puar.13105> (accessed 4 January 2022).
29. Nature Research and Greg Goodey, "State of Open Data Survey 2021 additional resources," figshare, Dataset, <http://doi.org/10.6084/m9.figshare.17081231.v1> (accessed 4 January 2022).
30. "QS World University Rankings 2021," QS Top Universities, 2021, <https://www.topuniversities.com/university-rankings/world-university-rankings/2021> (accessed 4 January 2022).
31. "Results and submissions," REF2021, 2022, <https://results2021.ref.ac.uk/filters/institution> (accessed 4 January 2022).
32. Herrmannova, Drahomira, Nancy Pontika, and Petr Knoth, "Do authors deposit on time? Tracking open access policy compliance," ACM/IEEE Joint Conference on Digital Libraries (JCDL) 2019, 206–216, DOI: <https://doi.org/10.1109/JCDL.2019.00037> (accessed 4 January 2022).
33. Ten Holter, "The repository, the researcher," 102079.
34. Beth Montague-Hellen and Kate Montague-Hellen, "Data Access Statement Analysis for UKRI funded papers."
35. "Policies and publication ethics," Frontiers, <https://www.frontiersin.org/guidelines/policies-and-publication-ethics> (accessed 5 January 2022).
36. "Data Availability," *PLOS ONE*, <https://journals.plos.org/plosone/s/data-availability> (accessed 5 January 2022).
37. "Data sharing," BMJ Author Hub, <https://authors.bmj.com/policies/data-sharing/> (accessed 5 January 2022).
38. "Research and Publication Ethics," MDPI, <https://www.mdpi.com/ethics> (accessed 5 January 2022).
39. "Reporting standards and availability of data, materials, code and protocols," *Nature Portfolio*, <https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards> (accessed 5 January 2022).
40. "Open Data Policy," *ASM Journals*, <https://journals.asm.org/open-data-policy> (accessed 5 January 2022).
41. "Data availability statement policy," OPTICA Publishing Group, <https://opg.optica.org/submit/review/data-availability-policy.cfm> (accessed 5 January 2022).
42. "IOP Publishing data availability policy," IOP Science, <https://publishingsupport.iopscience.iop.org/iop-publishing-data-availability-policy> (accessed 5 January 2022).
43. "Research Data," Oxford Academic, <https://academic.oup.com/pages/open-research/research-data> (accessed 5 January 2022).
44. "Data availability statements," *Springer Nature*, <https://www.springernature.com/gp/authors/research-data-policy/data-availability-statements> (accessed 12 January 2023).
45. "Publishing Policies and Ethics," EDP Sciences, <https://www.edpsciences.org/en/publishing-policies-ethics> (accessed 5 January 2022).
46. "Wiley's Data Sharing Policies," Wiley Author Services, <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html> (accessed 5 January 2022).
47. "BBSRC Safeguarding Good Scientific Practice," BBSRC, 2016, <https://www.ukri.org/wp-content/uploads/2022/01/BBSRC-260122-SafeGuardingGoodScientificPractice.pdf> (accessed 5 January 2022).
48. Mark Thorley, and Sarah Callahan, "NERC Data Policy – Guidance Notes," UKRI, 2019, <https://www.ukri.org/wp-content/uploads/2022/03/NERC-080322-PolicyGuidance-data-021219.pdf> (accessed 5 January 2022).
49. Frontiers, "Policies and publication ethics."
50. IOP Science, "IOP Publishing data availability policy."
51. BMJ Author Hub, "Data sharing."
52. Thorley and Callahan, "NERC Data Policy."

53. "Science Journals: editorial policies," *Science*, <https://www.science.org/content/page/science-journals-editorial-policies> (accessed 5 January 2022).
54. "Data Sharing," Royal Society of Chemistry, <https://www.rsc.org/journals-books-databases/author-and-reviewer-hub/authors-information/prepare-and-format/data-sharing> (accessed 5 January 2022).
55. "Data Repository Guidance," *Scientific Data*, <https://www.nature.com/sdata/policies/repositories> (accessed 5 January 2022).
56. "ACS Research Data Policy," ACS Publications, https://publish.acs.org/publish/data_policy (accessed 5 January 2022).
57. Frontiers, "Policies and publication ethics."

Article copyright: © 2023 Beth Montague-Hellen and Kate Montague-Hellen. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and distribution provided the original author and source are credited.



Corresponding author:
Beth Montague-Hellen
Head of Library and Information Services
The Francis Crick Institute
(Previously, Senior Research Librarian
University of Nottingham), UK
E-mail: beth.montague-hellen@crick.ac.uk
ORCID ID: <https://orcid.org/0000-0003-0946-1842>

Co-author:
Kate Montague-Hellen
ORCID ID: <https://orcid.org/0000-0002-8159-9013>

To cite this article:
Montague-Hellen B and Montague-Hellen K, "Publishers, funders and institutions: who is supporting UKRI-funded researchers to share data?," *Insights*, 2023, 36: 4, 1–17; DOI: <https://doi.org/10.1629/uksg.602>

Submitted on 06 July 2022

Accepted on 30 August 2022

Published on 14 February 2023

Published by UKSG in association with Ubiquity Press.