Reusable, FAIR humanities data: creating practical guidance for authors at Routledge Open Research

Rebecca Grant

F1000

Abstract

While stakeholders including funding agencies and academic publishers implement more stringent data sharing policies, challenges remain for researchers in the humanities who are increasingly prompted to share their research data.

This paper outlines some key challenges of research data sharing in the humanities, and identifies existing work which has been undertaken to explore these challenges. It describes the current landscape regarding publishers' research data sharing policies, and the impact which strong data policies can have, regardless of discipline.

Using *Routledge Open Research* as a case study, the development of a set of humanities-inclusive Open Data publisher data guidelines is then described. These include practical guidance in relation to data sharing for humanities authors, and a close alignment with the FAIR Data Principles.

Introduction

Reuse of research data is a core aim of stakeholders' research data policies, including those developed by academic publishers. As data policies become more stringent, are humanities authors being left behind?

In Spring 2022, Taylor & Francis and F1000 launched the *Routledge Open Research* publishing platform, aimed at researchers in Education, the Arts, Humanities, Psychology and Social Sciences. Open Data sharing is a key policy of the platform, and requires every author to share their data openly in a data repository and to draft a data availability statement describing how the data can be accessed. While there is evidence that strong publisher policies can advance research data sharing practices (Colavizza et al., 2020), humanities scholars are more likely than scientists to be limited by a lack of knowledge or awareness of open research, and may have concerns around issues such as IP and licensing (Longley Arthur & Hearn, 2021).

This paper describes the current landscape in relation to academic publishers' humanities research data sharing policies, and the challenges faced by humanities researchers when requested to share their research data. It then outlines the collaborative process undertaken by stakeholders from Routledge, Taylor & Francis and F1000[1] to develop a humanities-inclusive data sharing policy which provides practical guidance for authors and integrates emerging best practices from the broader humanities stakeholder community.

Humanities data and data sharing

Key research stakeholders are increasingly calling for the open publication of research data and the development of research data management plans to support data sharing. Funders including the European Commission,[2] National Institutes of Health (Kozlov, 2022), and the UKRI[3] mandate data sharing and evidence of data sharing, and many academic publishers provide research data sharing policies on their journals.[4]

As stakeholders' research data policies and mandates have increased in both stringency and prevalence since the 2010s, there has been some discussion around the relevance of "data sharing" for researchers in the humanities. Early projects such as KAPTUR (2011-2013), which explored the effective management of research data in the visual arts, identified challenges including data curators' ability to identify and classify the data, the heterogeneity of the data, and the use of research practices which can be "messy, fuzzy and tumultuous." (Garrett & Gramstadt, 2012).

Discussions on what constitutes humanities research data have continued without resolution, with the 2016 *State of Open Data* report maintaining that humanities researchers do not describe their research inputs as "data," and prefer terms including "sources" or "readings" (O'Donnell, 2016). Thoegersen (2018) notes that the reluctance of humanities researchers to identify their research outputs as data has directly impacted on how these outputs are managed. Acknowledging that the humanities have lagged behind in both data sharing and in adopting common frameworks and standards, the ALLEA e-humanities working group's "Recommendations for Sustainable and FAIR Data Sharing in the Humanities" opened for community consultation in June 2019,[5] with a final draft published by Harrower et al. in 2020.

---

[1] Routledge and F1000 are part of the Taylor & Francis Group.
[2] Data management: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm
[3] Publishing your research findings:
 https://www.ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-data-open/
[4] Center for Open Science TOP Guidelines: https://www.cos.io/initiatives/top-guidelines
[5] "E-Humanities," ALLEA All European Academies:
https://www.allea.org/working-groups/overview/working-group-e-humanities

As potential data identification, management and sharing practices have been discussed in the literature, the value of a common conceptualisation of research data in the humanities has also been explored. Peels & Bouter (2018) and Peels (2019) considered the potential benefits of publishing replicable humanities research, a practice which could be underpinned by preregistration of humanities studies, and sharing of data analysis plans, methods and datasets. Some humanities researchers have argued against this perspective, noting that it is neither possible nor desirable to replicate all humanities research, and that the diversity of interpretation is a strength of humanities research practice (Holbrook et al., 2019).

Other stakeholders have also addressed the intersection of humanities research and data sharing, for example the digital cultural heritage portal Europeana[6] and the humanities and social sciences data repository the Digital Repository of Ireland[7] have examined the FAIR Data Principles through a humanities lens. The EU research infrastructure DARIAH (the Digital Research Infrastructure for the Arts and Humanities) provides a rich "Pathfinder" resource for best practice data management in the humanities,[8] which links to training, guidance and toolkits. Funders in the humanities have also introduced data management and sharing policies, for example the Arts and Humanities Research Council (AHRC)[9] in the UK, and the Culture, Creativity and Inclusive Society strand of Horizon Europe.[10]

Over time, academic publishers have continued to implement data sharing policies across their journal portfolios, to introduce more consistency in data sharing requirements, and to enforce more stringent data sharing requirements. Through this process, new data sharing obligations are being introduced for humanities authors which do not necessarily reflect the range of perspectives, or challenges, associated with humanities data sharing.

Publisher data policies motivating change

In the past 15 years, academic publishers including Taylor & Francis,[11] F1000,[12] Wiley,[13] Elsevier[14] and Springer Nature[15] have introduced research data sharing policies which apply to authors submitting manuscripts to their journals or publishing platforms. Publisher policy frameworks tend to include a suite of policy options with common features, for example relating to deposition of data into repositories; submission of a data availability statement; data licensing; data citation; and data peer review. Depending on the policy selected for a particular journal, authors may be mandated to take action or are recommended to do so. In 2020 a set of common publisher data policy requirements was defined by Hrynaszkiewicz et al. via the

---

[6] Europeana and the FAIR principles for research data: https://pro.europeana.eu/post/europeana-and-the-fair-principles-for-research-data

[7] FAIR Principles: https://dri.ie/fair-principles

[8] DARIAH Pathfinder to Data Management Best Practices in the Humanities: https://campus.dariah.eu/resource/posts/dariah-pathfinder-to-data-management-best-practices-in-the-humanities

[9] Arts and Humanities Research Council (AHRC): https://www.dcc.ac.uk/guidance/policy/funders-data-policies/ahrc

[10] Cluster 2: Culture, Creativity and Inclusive society: https://ec.europa.eu/info/research-and-innovation/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/cluster-2-culture-creativity-and-inclusive-society_en

[11] Understanding our data sharing policies: https://authorservices.taylorandfrancis.com/data-sharing-policies

[12] Data Guidelines: https://f1000research.com/for-authors/data-guidelines

[13] Wiley's Data Sharing Policies: https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html

[14] Research Data Guidelines: https://www.elsevier.com/authors/tools-and-resources/research-data/data-guidelines

[15] Research Data Policy Types: https://www.springernature.com/gp/authors/research-data-policy/data-policy-types/12327096

Research Data Alliance, building on existing initiatives including the Transparency and Openness Promotion (TOP) Guidelines (Nosek et al., 2015) and allowing future publisher policy developments to follow a more consistent framework.

Subsequent research has examined the impact of journal data sharing policies, including the increased prevalence of data availability statements due to strengthened data policies at PLOS and BMC journals (Colavizza et al., 2020), and at Wiley (Graf et al., 2020). Positive correlations have also been demonstrated regarding data sharing and increased rates of article citation (Christensen et al., 2019 and Colavizza et al., 2020). Journal data sharing policies have the potential to change researchers' practices, while also providing tangible benefits in the form of increased article citations. There is also evidence that additional data sharing requirements from publishers do not negatively impact on submission or acceptance rates at their journals (Cannon et al., 2022).

Although the implementation of research data policies at publishers has increased, and research papers on the impacts of such policies are being produced, little has been written about the impact of data sharing policies at humanities journals. Few humanities journals have research data policies which mandate that authors share their data openly in data repositories, with exceptions including *Humanities & Social Sciences Communications*,[16] and *Emerald Open Research* (an F1000 platform);[17] and data journals including *The Journal of Open Archaeology Data*,[18] *The Journal of Open Humanities Data*,[19] and *Research Data Journal for the Humanities and Social Sciences*.[20]

In 2020 Springer Nature announced that its Springer imprint journals (including those in the humanities) would move to a policy which requires authors to include a data availability statement with their manuscript (Grant, 2020). While the Springer policies (drafted by this author) include some wording specific to humanities research, generally speaking the language of publisher research data policies aimed at humanities researchers is indistinguishable from those written for biologists, chemists or physicists. Given that a key issue associated with data sharing in the humanities is a lack of understanding of what constitutes data, this lack of acknowledgement of humanities research practice constitutes a "language barrier" for humanities authors.

Publishers and humanities data sharing: the STM Humanities working group

The STM Association is a global trade association for academic and professional publishers, with 145 members in 21 countries.[21] The Association declared 2020 to be its Research Data Year,[22] underpinned by objectives which supported its members to introduce research data policies which encouraged data availability statements, data sharing and data citation (Smit & van Rossum, 2022). This work continued in 2021 and 2022, leading to the creation of a Humanities Data focused sub-group.[23] Representatives from publishers including F1000, Taylor & Francis, Wiley, SAGE, Brill, Cambridge University Press and Oxford University Press use this

---

[16] Editorial and publishing policies: https://www.nature.com/palcomms/journal-policies/editorial-and-publishing-policies#Availability%20of%20materials%20and%20data

[17] Data Guidelines: https://emeraldopenresearch.com/for-authors/data-guidelines

[18] About this Journal: https://openarchaeologydata.metajnl.com/

[19] Editorial policies: https://openhumanitiesdata.metajnl.com/about/editorialpolicies

[20] Research Data Journal for Humanities and Social Sciences: https://brill.com/view/journals/rdj/rdj-overview.xml

[21] STM Association: https://www.stm-assoc.org

[22] STM 2020 Research Data Year: https://www.stm-assoc.org/standards-technology/2020-stm-research-data-year

[23] STM Research Data Program - Launch of Humanities-Focused Subgroup: https://www.stm-assoc.org/humanitiesresearchdata

group as a space to discuss the specific challenges of research data policies for humanities journals, editors and authors. In 2022, its work includes the development of a survey of authors in the humanities and interviews with editors of humanities journals, with the aim of gathering concrete data related to the challenges and opportunities of research data policies at humanities journals.

Some preliminary findings from this survey are presented below, and the full anonymised dataset and an analysis of the results will be published by the STM's Humanities Data sub-group in 2022. Figure 1 illustrates the extent to which respondents recognised the term "research data" when considering the outputs of their research process. Over 80% of the respondents did recognise this term as being relevant to their research at least some of the time, although this reflects the self-selecting group of respondents: humanities authors who chose to respond to a survey specifically relating to research data. Previous research indicates that many humanities researchers do not prefer to use this terminology (O'Donnell, 2016).
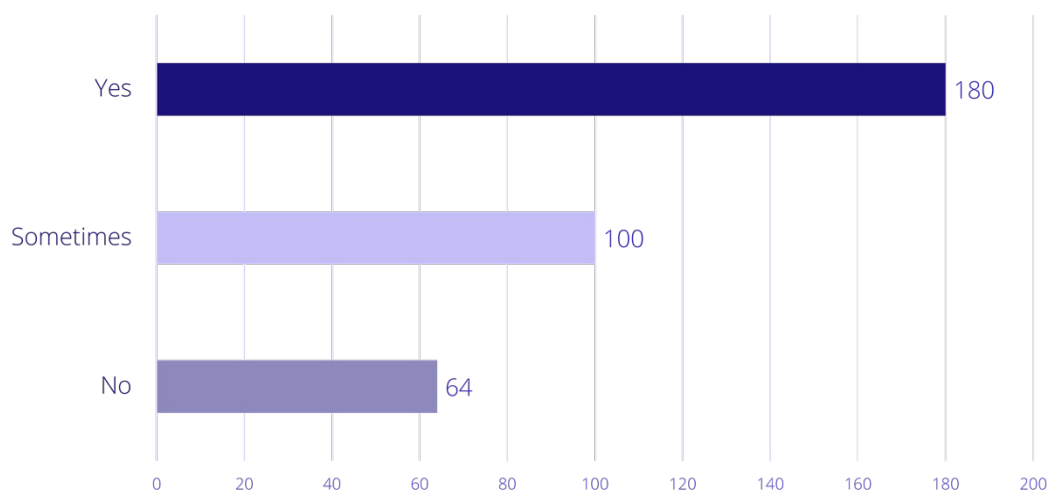

Figure 1.   Survey responses to the question "Do you recognise the term 'research data' as a supporting output of your research process?"

Respondents were also asked whether they believed that it was important for research data to be shared with others in the humanities (Figure 2) and nearly 89% believed that it was either very important, or important. However, when asked whether they had shared their research data, and how, the majority (217 respondents) reported that they shared research data using a personal, peer-to-peer method such as email (Figure 3). Less than half as many (104) had shared by uploading their data to a repository, which is the method best aligned with Open Data sharing policies from funders and publishers, and the FAIR Data Principles. Only 17 stated that they did not share their research data as they did not want to (as opposed to more concrete reasons such as lack of permission to share).
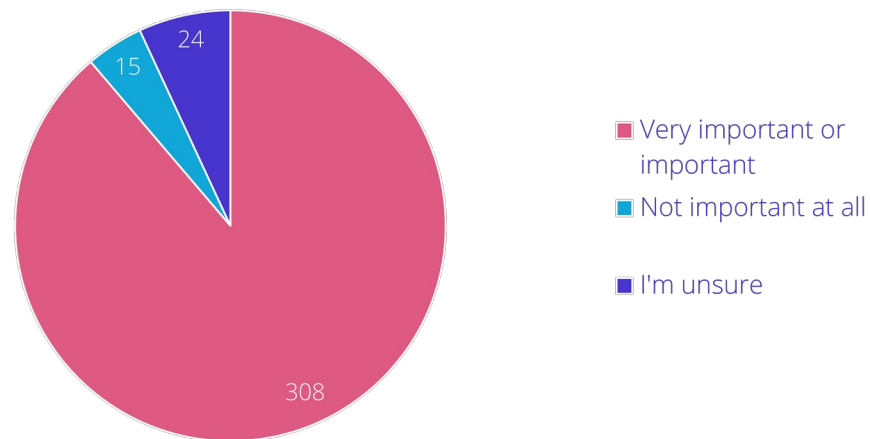
Figure 2. Survey responses to the question "How important do you think it is to share research data with others in the humanities?"
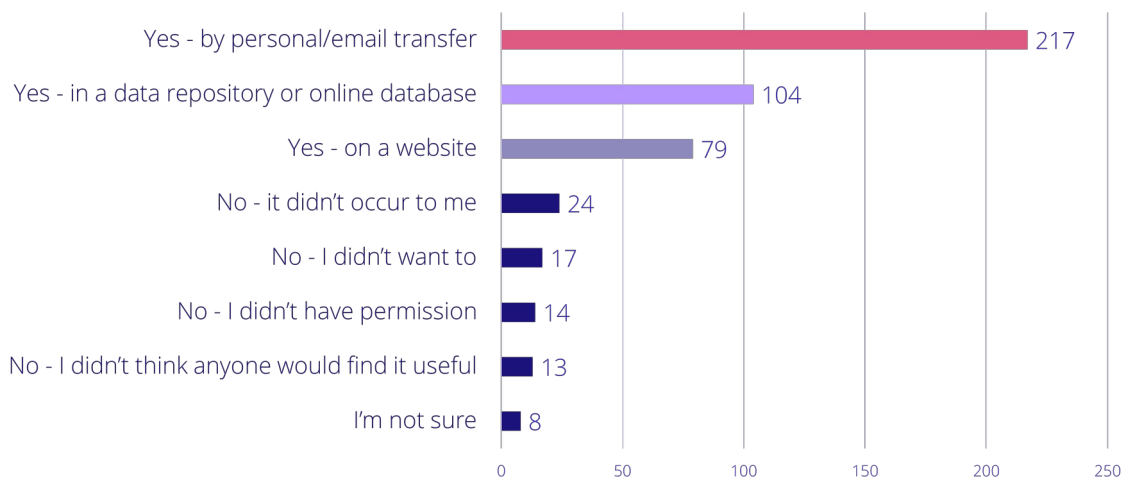

Figure 3. Survey responses to the question "Have you ever shared your research data with others?"

These survey responses, although limited by the self-selecting pool of respondents, indicate that there is an understanding and enthusiasm from a proportion of humanities researchers to share their research data. However, when asked about their sharing methods, many researchers are not aligning with FAIR and do not deposit their data into a data repository, instead relying on peer-to-peer sharing or project websites. These preliminary results demonstrate a willingness and openness to sharing research data by humanities researchers, which can be leveraged through the creation of relevant, targeted guidance which addresses definitions of humanities "research data" and provides guidance on sharing FAIR, Open Data.

Research data policies for humanities and social sciences authors: *Routledge Open Research*

Routledge Open Research is a publishing platform which combines F1000's innovative publishing model and technology with Routledge's global reputation and reach, providing the first dedicated open research platform for arts, humanities and social sciences scholars.[24] The F1000 model is based on an open and transparent publishing process where manuscripts are published rapidly following editorial checks; peer review occurs openly with all revisions made public; and all data associated with the study must be deposited to an appropriate data repository with an open licence.[25]

At launch, *Routledge Open Research* was intended to enforce the same Open Data sharing policy which is in place on other F1000 platforms.[26] The F1000 Open Data policy is supported by extensive guidelines which outline the process of sharing FAIR and Open Data, from creation to publication;[27] however like other publishers' data policies, these guidelines are aimed at a multi-disciplinary audience and do not account for the specific data sharing practices of humanities researchers.

Ahead of the launch of the platform, the Open Data guidelines for *Routledge Open Research* were redrafted with the following aims:

I. To distil the guidelines into a list of clearly defined steps for authors to take.

II. To ensure that the guidelines were appropriate and understandable regardless of whether the authors generated new research data, reused existing research data, or based their research on other materials or sources such as archives or museum collections.

III. To integrate any relevant external guidance that would support humanities authors to share their data.

Although the guidelines were drafted to support both humanities researchers and social scientists, this paper focuses on the humanities perspective.

Developing humanities-inclusive data guidelines

The development of the *Routledge Open Research* data guidelines was a collaborative process. Led by F1000's Head of Data and Software Publishing, an initial set of requirements for the guidelines were established in a workshop with F1000 editorial and Taylor & Francis Open Research colleagues. A draft set of guidelines was then created which underwent several rounds of revision, incorporating the perspectives of the Open Research, editorial and publishing teams from F1000, Taylor & Francis and Routledge. Suggestions on relevant external guidance were solicited from these teams, and via the STM Association's Humanities group. Once published, continued feedback from authors and advisors at *Routledge Open Research* will be considered and integrated where relevant.

---

[24] An Introduction to the new Routledge Open Research Platform:
https://think.f1000.com/routledgeopenresearch/
[25] Why F1000?: https://f1000.com/why-f1000
[26] Data availability: https://f1000research.com/about/policies#dataavail
[27] Data Guidelines: https://f1000research.com/for-authors/data-guidelines

Distilling the Open Data guidelines into a list of clearly defined steps

While F1000's existing Open Data guidelines provide extensive detail on all aspects of data sharing, they assume that the author will start their process early in the research lifecycle, potentially before the research has begun or any data has been collected. Firstly, the FAIR Data Principles are explained to the reader, and guidance for preparing spreadsheets for sharing is given; this is followed by an extensive list of recommended repositories from all disciplines, and subsequently additional information on data availability statements, data citation and data linking.

To create the new guidelines, each mandatory requirement was separated from the body of the text and used to create a new numbered list of an author's obligations when publishing on *Routledge Open Research*. They are drafted with the assumption that the researcher will only encounter them at the point where they are ready to submit a research article, rather than at the beginning of their research project. The draft of the list was as follows:

1. Your dataset(s) must be deposited in an appropriate data repository.

2. Your dataset(s) must have a license applied which allows reuse by others (CC0 or CC-BY).

3. Your dataset(s) must have a persistent identifier (e.g., a DOI), allocated by a data repository.

4. You must provide a data availability statement as a section at the end of your article, including elements 1-3.

5. You must include a data citation and add a reference to data to your reference list.

6. Your dataset(s) should not contain any sensitive information, for example in relation to human research participants.

7. You should share any related software and code.

8. Your dataset(s) must be useful and reusable by others, adhere to any relevant data sharing standards in your discipline and align with the FAIR Data Principles.

9. Your dataset(s) should link back to your article, if possible.

This list of requirements forms the introduction to the new guidelines, with additional information on each point provided in the sections which follow them. The intention is that authors who are more familiar with data sharing can see the mandatory requirements at a glance, while others can access more information if needed.

Ensuring the guidelines are appropriate for all humanities authors

Although the mandatory aspects of the guidelines had been clarified, they still had the potential to confuse authors who had reused existing datasets, or those who had worked with physical or digital sources including heritage collections from galleries, libraries, archives and museums (GLAM institutions). In those cases, an author would not have had the opportunity to deposit the dataset underpinning their research into a data repository, to apply an open licence to it, or to obtain a persistent identifier. Some additional guidance was therefore needed, as *Routledge Open Research* will accept manuscripts based on these types of data, as long as the sources are clearly described in an accompanying data availability statement.

The list of mandatory requirements was extended, to encompass all possible scenarios, including the "Open Data requirements for new datasets you have generated" listed above. Two additional, shorter lists of requirements were created to address authors who had reused

existing, third-party datasets, and for authors who had used other types of "data" including sources from museums, archives, and gallery collections, whether physical or digital:

Open Data requirements for existing datasets which you have reused:

1. You must provide a data availability statement as a section within your article, describing where the dataset is located and how it can be accessed by others.

2. You must include a citation for each dataset and add it to your reference list.

Accessible Source Materials requirements:

1. You must provide a data availability statement as a section within your article, describing where the source material is located and how it can be accessed by others.

2. You must include a citation for your source and add it to your reference list.

The section on Accessible Source Materials relates to humanities data sources including heritage collections and provides additional guidance on how authors can present these sources in data availability statements and citations. As an author will not have control over licences, persistent identifiers or the location of these resources, the guidelines make it clear that transparency is all that is required. The data availability statement should simply describe how the materials were accessed, and how others might access them in future.

Integrating relevant external guidance

To avoid "reinventing the wheel" or replicating work previously done by other stakeholders in this space, efforts were made to identify and integrate any existing guidance or initiatives which might help humanities authors to better understand the data sharing requirements. Several resources were linked from the new guidelines, including:

1. "The Heritage Data Reuse Charter: from principles to research workflows" (Tóth-Czifra & Romary, 2020) to assist humanities researchers in considering how data sharing applies to their research practice and the reuse of data from cultural heritage institutions.

2. "Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities," (Harrower et al., 2020) to provide additional guidance on sharing FAIR humanities data.

3. The Heritage PIDs project website[28] for guidance on data citation for digitised cultural heritage collections.

The inclusion of these resources is intended to provide supplementary contextual information which authors can choose to explore, but it is not necessary to read them in order to understand or comply with the Open Data policy.

Providing relevant examples

Given that the guidelines are intended for authors who have not previously shared their research data, it was important to include illustrative examples of appropriate data availability statements and data citations (Table 1). These were selected from existing published articles on F1000

---

[28] How to provide recommended citations for cultural heritage items?:
https://tanc-ahrc.github.io/PIDResources/Citation%20guidance.html

platforms, including *Gates Open Research*, *Emerald Open Research*, and *Wellcome Open Research*, and include examples from authors who have used data including census records, third-party survey data, and physical cultural heritage collections. Other examples are available on the *Routledge Open Research* data guidelines webpage.

Data type	Example Data Availability statement	Example citation
Census data	Individual-level data from the 1940 US Census is available from IPUMS https://doi.org/10.18128/D010.V8.0.EXT1940USCB8. These data are under Copyright of Minnesota Population Center, University of Minnesota. Access to the documentation is freely available without restriction; however, users must register before extracting data from the website. The output of the TopDown algorithm when run on the 1940 US Census data is available to download from the US Census Bureau: https://www2.census.gov/census_1940/. These data are under Copyright of the United States Census Bureau.	Ruggles S, Flood S, Goeken R, et al.: IPUMS USA: Version 8.0 extract of 1940 Census for U.S. Census Bureau disclosure avoidance research [dataset]. 2018. http://www.doi.org/10.18128/D010.V8.0.EXT1940USCB Example taken from: Petti S and Flaxman AD. Differential privacy in the 2020 US census: what will it do? Quantifying the accuracy/privacy tradeoff. Gates Open Res 2020, 3:1722 (https://doi.org/10.12688/gatesopenres.13089.2)
Third-party survey data	DCMS Survey. One dataset used in this paper is composed of responses to DCMS survey release for quarter 4 of their longitudinal survey titled 'Taking Part' (see DCMS (2017a), available at: https://www.gov.uk/government/statistics/taking-part-201617-quarter-4-statistical-release (accessed 30-Jan-2020) under the Open Government License agreement. BFI survey. One dataset used in this paper is composed of responses to a British Film Institute (BFI) survey titled: 'Cultural Consumption' conducted by IpsosMORI in 2011. The BFI provide the survey dataset as appendix 4 of their larger report: Northern Alliance and Ipsos MediaCT (2011) Opening our eyes: How film contributes to the culture of the UK (Report), London: British Film Institute. Available at: https://www.bfi.org.uk/about-bfi/policy-strategy/opening-our-eyes-how-film-contributes-culture-uk	DCMS: Taking Part Survey: England Adult Report, 2016/17 – Statistical release quarter 4' [dataset]. DCMS/Crown: London. 2017a; (accessed 30-Jan-2020). https://www.gov.uk/government/statistics/taking-part-201617-quarter-4-statistical-release. *Example taken from: Hanchard M, Merrington P, Wessels B et al. Developing a computational ontology to understand the relational aspects of audience formation. Emerald Open Res 2020, 2:5 (https://doi.org/10.35241/emeraldopenres.13465.1)*
Data from an archival collection	The collection used for this paper, The Papers of Victor Webb, is held by University of Glasgow Archives and Special Collections. The catalogue is available online: https://archiveshub.jisc.ac.uk/search/archives/e81a4437-f0b2-33ad-8229-ef4fc8a0cfdf.	Example taken from: Connelly H. A place to grow: Well-being and activism on Edinburgh's post-war allotments and how this can inform urban gardening in Scotland today]. Wellcome

Data type	Example Data Availability statement	Example citation
	You can access the papers by making an appointment in the Searchroom here: https://www.gla.ac.uk/myglasgow/archives/contact/searchroombookingform/.	Open Res 2019, 4:72 (https://doi.org/10.12688/wellcomeopenres.15216.1)
	A small number of records in this collection are subject to Data Protection legislation as they contain sensitive information; however, access may be given to bona fide researchers and academics. Please contact the Duty Archivist for advice on how to apply for access to these files (enquiries@archives.gla.ac.uk).	

Table 1. Examples of appropriate data availability statements and data citations for various data types.

The examples provided will be updated to include additional data types, as appropriate examples are published on the *Routledge Open Research* platform.

Alignment with the FAIR Data Principles

Like other F1000 platforms, *Routledge Open Research* endorses the FAIR Data Principles, and its author guidelines provide practical suggestions to help authors to create and share FAIR research data. Across all disciplines, researchers may find elements of the FAIR principles challenging to implement. David et al. (2020) note that the Principles are not easily put in practice by researchers, and are also difficult to interpret by evaluators. They may be even more challenging for humanities authors, due to the physical nature of some of their sources or data, the ownership or copyright attached to the data, and the lack of community-level agreement on what constitutes the data (Tóth-Czifra, 2020). The ALLEA report "Recommendations for Sustainable and FAIR Data Sharing in the Humanities" (Harrower et al., 2020) addresses some of these challenges, covering topics including identifying research data, addressing legal issues, and considerations around file formats, metadata and data models, and provides recommendations to humanities researchers for each.

Rather than providing guidance for researchers which would be applicable from the beginning of their research project, the *Routledge Open Research* guidelines map the individual elements of FAIR to a series of data sharing steps which can be undertaken ahead of article submission, including choosing a repository, drafting metadata, and applying an open licence (Table 2). This guidance is supplemented by the provision of a list of recommended repositories for authors to choose from, information on open licensing, and a recommendation to consult FAIRsharing.org. This is intended to address the potential barrier of technical language used in the FAIR Principles themselves (for example "standardised communications protocol") and to provide suggestions that can be actioned relatively quickly as the author prepares to submit their article.

FAIR acronym	Associated Principles	Inclusion in *Routledge Open Research* guidance
Findable	F1. (Meta)data are assigned a globally unique and persistent identifier. F2. Data are described with rich metadata (defined by R1 below). F3. Metadata clearly and explicitly include the identifier of the data they describe. F4. (Meta)data are registered or indexed in a searchable resource.	A list of recommended data repositories is provided for authors; each repository has been assessed to ensure that it assigns unique persistent identifiers, and that its contents are indexed and searchable. Additionally authors are encouraged to include as much contextual information (metadata) as possible when depositing datasets to their chosen repository.
Accessible	A1. (Meta)data are retrievable by their identifier using a standardised communications protocol A1.1 The protocol is open, free, and universally implementable A1.2 The protocol allows for an authentication and authorisation procedure, where necessary A2. Metadata are accessible, even when the data are no longer available	All repositories on the *Routledge Open Research* recommended repositories list use standard communications protocols (like http://), and have been checked to ensure that they will continue to provide access to metadata even when datasets are removed.
Interoperable	I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (Meta)data use vocabularies that follow FAIR principles I3. (Meta)data include qualified references to other (meta)data	Authors are encouraged to check the website FAIRsharing.org to establish whether any applicable standards are available, and to use them if so; as well as using open, non-proprietary file formats for your data. Additionally, the *Routledge Open Research* recommended data repositories are assessed to ensure that they allow links or references to related datasets.
Reusable	R1. (Meta)data are richly described with a plurality of accurate and relevant attributes R1.1. (Meta)data are released with a clear and accessible data usage license R1.2. (Meta)data are associated with detailed provenance R1.3. (Meta)data meet domain-relevant community standards	The guidelines reiterate the recommendation that authors add as much contextual information (metadata) as possible when depositing the dataset into a repository, and that authors should check the website FAIRsharing.org to establish whether any applicable standards are available, and to use them if so. Additionally, authors are required

FAIR acronym	Associated Principles	Inclusion in *Routledge Open Research* guidance
		to apply an open license (CC-BY) or Public Domain Dedication (CC0) to their dataset in the repository.

Table 2. The FAIR Data Principles and associated guidance in the *Routledge Open Research* data guidelines.

Next steps and Conclusion

The *Routledge Open Research* platform was launched in April 2022, and the data guidelines developed for this platform will be made available on all F1000 platforms publishing humanities and social sciences articles in future. *Routledge Open Research* is also intended to support Open Plus Books,[29] allowing authors to rapidly publish books and book chapters. This will provide the opportunity to explore data availability for book authors, and potentially necessitate the update or adaptation of the guidelines. In the meantime, compliance with the *Routledge Open Research* Open Data policy and guidelines is expected to be 100%, as authors cannot continue from submission to publication without addressing their requirements. Any emerging issues and feedback will be captured by the platform's pre-publications team, who will also be available to provide additional support to authors who need it.

The development of the Open Data guidelines for *Routledge Open Research* represents the creation of a unique resource for humanities researchers. While the principles of FAIR data and reusability are applicable across research disciplines, they have not previously been presented as a set of practical steps for humanities authors to take when publishing in an academic journal. We hope to gather feedback from authors, advisors and the editorial team at *Routledge Open Research*, as well as other stakeholders in the community, and continue to improve the guidelines and make them even more useful over time. Additionally, we hope that these practical guidelines will encourage other publishers, and stakeholders such as funding agencies, to continue to develop their guidance to support data sharing by humanities researchers.

Acknowledgements

---

[29] Open Plus Books: Evolving knowledge for research communities: https://www.routledge.com/our-products/open-access-books/open-plus-books

References

[journal article] Cannon, M., Kelly, A., & Freeman, C. (2022). Implementing an Open & FAIR data sharing policy—A case study in the earth and environmental sciences. *Learned Publishing, 35, 1*, (pp. 56-66). Doi: https://doi.org/10.1002/leap.1442.

[journal article] Christensen, G., Dafoe, A., Miguel, E., Moore, D.A., Rose, A.K. (2019). A study of the impact of data sharing on article citations using journal policies as a natural experiment. *PLoS ONE 14(12): e0225883*. Doi: https://doi.org/10.1371/journal.pone.0225883.

[journal article] Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLoS ONE 15(4): e0230416*. Doi: https://doi.org/10.1371/journal.pone.0230416.

[journal article] David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., Jonquet, C., Dollé, L., Jacob, D., Bailo, D., Bravo, E., Gachet, S., Gunderman, H., Hollebecq, J.-E., Ioannidis, V., Le Bras, Y., Lerigoleur, E., Cambon-Thomsen, A. and the SHAring Reward and Credit (SHARC) Interest Group (2020). FAIRness Literacy: The Achilles' Heel of Applying FAIR Principles. *Data Science Journal*, 19(1), p.32. DOI: http://doi.org/10.5334/dsj-2020-032.

[journal article] Garrett, L. & Gramstadt, M.-T. (2012). KAPTUR: exploring the nature of visual arts research data and its effective management. *EVA LONDON 2012 Electronic Workshops in Computing*, 88-96. Doi: 10.14236/ewic/EVA2012.16.

[journal article] Graf, C., Flanagan, D., Wylie, L. & Silver, D. (2020). The Open Data Challenge: An Analysis of 124,000 Data Availability Statements and an Ironic Lesson about Data Management Plans. *Data Intelligence, 2, 4,* (pp. 554–568). Doi: https://doi.org/10.1162/dint_a_00061.

[blogpost] Grant, R. (2020, March 23) What does data sharing mean for the humanities and social sciences? Retrieved from https://www.springernature.com/gp/researchers/the-source/blog/blogposts-research-integrity/what-does-data-sharing-mean-for-humanities-and-social-sciences/18982506.

[report] Harrower, N., Maryl, M., Biro, T., Immenhauser, B., & ALLEA Working Group E-Humanities. (2020). *Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities*. Doi: https://doi.org/10.7486/DRI.tq582c863.

[blogpost] Holbrook, J. B., Penders, B. & de Rijcke, S. (2019, January 21). The humanities do not need a replication drive. Retrieved from https://www.cwts.nl/blog?article=n-r2v2a4&title=the-humanities-do-not-need-a-replication-drive.

[journal article] Hrynaszkiewicz, I., Simons, N., Hussain, A., Grant, R. and Goudie, S. (2020). Developing a Research Data Policy Framework for All Journals and Publishers. *Data Science Journal*, 19, 1, (p. 5). doi: http://doi.org/10.5334/dsj-2020-005.

[journal article] Kozlov, M. (2022). NIH issues a seismic mandate: share data publicly. *Nature, 602*, 558-559. Doi: https://doi.org/10.1038/d41586-022-00402-1.

[journal article] Longley Arthur, P. & Hearn, L. (2021). Toward Open Research: A Narrative Review of the Challenges and Opportunities for Open Humanities. *Journal of Communication, 71, 5,* (pp. 827–853). Doi: https://doi.org/10.1093/joc/jqab028.

[journal article] Nosek, B. A. et al. (2015). Promoting an open research culture. *Science, 348, 6242*, (pp. 1422-1425). Doi: 10.1126/science.aab2374.

[report] O'Donnell, D. (2016). The Bird in Hand: Humanities Research in the Age of Open Data. In Treadway, J., Hahnel, M., Leonelli, S., Penny, D., Groenewegen, D., Miyairi, N., Hayashi, K., O'Donnell, D., Science, D., & Hook, D., *The State of Open Data.* Doi: https://doi.org/10.6084/m9.figshare.4036398.v1.

[journal article] Peels, R. (2019). Replicability and replication in the humanities. *Research Integrity and Peer Review 4, 2*. Doi: https://doi.org/10.1186/s41073-018-0060-4.

[journal article] Peels, R., & Bouter, L. (2018) The possibility and desirability of replication in the humanities. *Palgrave Communications, 4, 95*. Doi: https://doi.org/10.1057/s41599-018-0149-x.

[journal article] Smit, E. & van Rossum, J. (2022). STM's Research Data Year: Results, lessons learnt, and next steps. *Learned Publishing, 35* (pp. 67-70). Doi: https://doi.org/10.1002/leap.1434.

[journal article] Thoegersen, J. L. (2018). 'Yeah, I Guess That's Data': Data Practices and Conceptions among Humanities Faculty. *Libraries and the Academy 18, 3,* 491-504.

[report] Tóth-Czifra, E. & Romary, L. (2020). *The Heritage Data Reuse Charter: from principles to research workflows*. Retrieved from https://halshs.archives-ouvertes.fr/halshs-02475692.

[book chapter] Tóth-Czifra, E. (2020). The Risk of Losing the *Thick Description*: Data Management Challenges Faced by the Arts and Humanities in the Evolving FAIR Data Ecosystem. In J. Edmond (Ed.), *Digital Technology and the Practices of Humanities Research* (pp. 235-263). Open Book Publishers. Doi: https://doi.org/10.11647/OBP.0192.