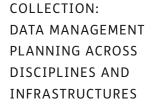# Supporting FAIR Data Management Planning Across Different Disciplines at the University of Sheffield

JENNI ADAMS ⓘD

BEV JONES ⓘD

HELEN FOSTER

*Author affiliations can be found in the back matter of this article

ABSTRACT

Recognition is growing sector-wide of the importance of FAIR data management planning in facilitating the sharing and reuse of research outputs. Indeed, research funders increasingly mandate practices such as data sharing or inclusion of a data availability statement in publications issuing from funded projects, practices which need to be anticipated at the data management planning stage. Nevertheless, there is a shortage of discipline-specific guidance to support researchers seeking to make data and software FAIR within a given domain and to a meaningful degree. This article outlines a project which explored ways to address this shortage. In Spring 2022, the University of Sheffield Library worked with researchers in seven disciplines to develop subject-specific FAIR checklists for the use of colleagues before, during and at the end of their research project. We chart the outcomes and implications of the project and the ways this will inform future support for researchers in planning for FAIR data at this institution.

INTRODUCTION

The acronym FAIR refers to Findable, Accessible, Interoperable and Reusable research data and software, as embodied in the FAIR data principles (Wilkinson et al 2016) and, more recently, FAIR4RS (Chue Hong et al. 2022). To give a few examples, practices that increase findability and accessibility include publishing in a repository which issues a DOI and ensuring the data or code is fully documented; practices that increase interoperability and reusability include providing as complete metadata as possible, using standard or open file formats, creating a README file, and using a Creative Commons licence.

As we detail below, shaping one's data management plan (DMP) around FAIR practices creates multiple benefits, yet while a number of useful resources have recently been produced,[1] there remains a shortage of guidance on how best to achieve FAIR within a given discipline. In this article, we outline a recent initiative at the University of Sheffield (TUoS) to explore potential methodologies and sticking points in achieving this. A pilot project saw staff from the University Library's Research Data Management (RDM) team work with seven departments covering all faculties of the university, employing a small team of postgraduate and early career researchers in each department to create a FAIR data and software checklist for their discipline. These small teams of 'Department FAIR leads' surveyed colleagues on their existing understanding and uptake of FAIR practices before using workshops to develop and finesse guidelines. As a related initiative, colleagues in the University Library worked separately with a number of researchers across different faculties to develop a series of video case studies highlighting existing good practice around FAIR (Adams 2022b). This article will focus on the checklist project, since this most clearly relates to guidance to support data management planning.

THE BENEFITS OF FAIR DATA AND SOFTWARE

Before unpacking the project, it would be useful to reflect on why it is so crucial that researchers integrate FAIR practices into their DMP. Making research data and software FAIR enables verifiability of methods, analyses and conclusions, allowing greater corroboration of a study's academic rigour (Tenopir et al. 2011; Fecher et al. 2015; Munafò 2017; Norris & O'Connor 2019); furthermore, making methods and outputs FAIR benefits the researcher by ensuring continuing accessibility and reusability when data is revisited (McKiernan et al. 2016). Sharing data and software by applying FAIR principles also promotes collaboration and academic progress, enabling researchers to avoid duplicating existing work and to develop innovative ways to build on existing materials (Pierce et al. 2019; Tenopir et al. 2011; McKiernan et al. 2016). From the perspective of value for money, FAIR practices also ensure that the (often public) funds which support research yield an increased impact with greater longevity (Tenopir et al. 2011).

In order to achieve these benefits fully and efficiently, FAIR practices should ideally be formulated at the DMP stage as this ensures that all necessary mechanisms are present and in alignment. For instance, a researcher wishing to share de-identified data via a repository using a CC-BY licence will need to first ensure that appropriate consent is gathered, that their de-identification process is carefully planned, that a consistent file structure is used and that any necessary permissions are obtained. Nevertheless, it should be acknowledged that data and software can be FAIRified at a later stage, albeit less straightforwardly (for example, rationalising file storage after data collection, or obtaining retrospective consent for sharing). Given that many researchers are only now becoming aware of the FAIR principles, in many cases these retrospective applications of FAIR principles will be necessary. Guidance about FAIR must thus be applicable both to researchers at the DMP stage and to those applying the principles retrospectively.

OBSTACLES TO FAIR

Given the overwhelming benefits of FAIR data and software, why are FAIR practices not already universally adopted? One answer is that there are significant time and workload implications to doing so (Tenopir et al. 2011; Huang et al. 2012; Fecher et al. 2015; Abele-Brehm et al.

2019; Perrier et al. 2020; Zhu 2020).[2] Making data and/or software FAIR creates additional considerations at each stage, from more complex and detailed data management planning, through participant consent, data organisation, data preparation, creating metadata and README documentation, and so on. Especially where these processes need to be applied retrospectively, the task may appear overwhelming and may be perceived by overworked researchers as merely an additional source of stress. This is compounded by what many regard as a lack of incentivisation to share data and software (Perrier et al. 2020; Fecher et al. 2015; Stewart et al. 2021).

Turning from generic to discipline-specific objections, researchers may find it especially challenging to apply FAIR practices in disciplines (traditionally, in Arts and Humanities) where the concept has less existing traction (Calamai & Frontini 2018). In some disciplines, fewer research projects may be substantially funded (Suber 2016), resulting in discipline-specific resourcing issues. Concerns about the appropriateness and/or applicability of FAIR principles to an individual discipline or study also exist. Zhu (2020) suggests that Arts and Humanities researchers often do not identify 'data' as a relevant concept to their study. In Social Sciences and Health-related research, there may be concerns about participants' privacy, with the necessary culture of data protection creating resistance to FAIR and in some cases a misapprehension of the concept: FAIR principles advocate for data and software to be 'as open as possible, as closed as necessary' (H2020 Programme 2016) rather than mandating unrestricted sharing.

These obstacles can be substantially addressed by advocacy and guidance around FAIR which accommodates discipline-specific issues, practices and concerns, in addition to highlighting the achievability and importance of factoring FAIR practices into data management planning. At TUoS, our FAIR roadmap brings together a number of workflows and projects to address these aims as part of the stated goal to 'Develop support and training for FAIR data practices … , recognising the diverse requirements across disciplines and ensuring appropriate long-term preservation of, and access to, research data', a priority in the University's Research Strategy Delivery Plan (University of Sheffield 2021). The project under discussion here formed one aspect of the roadmap and aimed to raise the profile of FAIR principles among researchers at TUoS while providing practical advice on how to integrate the principles into specific research contexts.

THE PROJECT

Using Research England funds designated for research culture change, the RDM team in the TUoS Library formulated a pilot project to investigate potential ways to develop discipline-specific guidance on FAIR to inform data management planning and retroactive FAIRification in the departments involved. The project took place between April and July 2022, with the short timescale being a requirement of the funding. Seven departments covering all faculties were selected, with chosen departments providing a broad spread of disciplines. The departments selected for the pilot project were Architecture (People, Environments and Performance group); Geography; Biosciences (Ecology and Evolutionary Biology and Molecular and Cellular Biology groups); Psychology; Mechanical Engineering; English and Health and Related Research. The decision to focus in some cases on a whole department and in others on a specific research group resulted from discussions with departmental Directors of Research and department leads. In some cases, the research practices and data types used within a department varied so substantially that attempting to accommodate all practices and recommendations within a single checklist would have been unachievable.

Once departmental agreement was secured, either one or two 'Department Leads' (postgraduate and early career researchers in pilot departments) were appointed for each department and given detailed training on RDM, FAIR, FAIR4RS and the project methodology. Department leads were each supported by a colleague in the Library's RDM team. In each department, a survey to gauge the existing awareness and take-up of FAIR practices among staff and PGRs and to gather examples of good practice was followed by a workshop for interested colleagues to further discuss their suggestions and recommendations as well as resources identified

---

2   While these researchers frame their analysis in terms of 'data sharing' rather than FAIR, the large degree of overlap means that their findings are also applicable to this context.

by departmental leads. Suggested topics for workshop discussion included the benefits and challenges of FAIR in the specific disciplinary context; departmental researchers' concerns around FAIR; instances of existing good practice; feedback on the resources identified by departmental leads; and thoughts on the format of resource colleagues would find useful.

Drawing on the information gathered through surveys, workshops and their own research, and in consultation with RDM colleagues, department leads then created drafts of their checklists. As a starting point, they were provided with a template from the University Library containing general guidance on making data and software FAIR. Department leads developed and customised this, adding guidance on topics including discipne- and data-type-specific repositories, optimal data formats, and advice on discipline-specific issues in data sharing such as restricted access to sensitive health data. Department leads had been briefed on the need to find a balance between brevity and depth that was fitting to the needs of colleagues, the value of incorporating links to broader information, and the need to consider the checklist's structure and scope. They were also encouraged to book onto a 'Code Clinic' – a short consultation with colleagues in the University's Research Software Engineering team (see University of Sheffield n.d.) – in order to sense-check their guidance on research software. Through this process and via consultation with interested colleagues, department leads finessed their checklist before presenting it to disciplinary colleagues at a launch event.

OBSERVATIONS

Feedback gathered during the project indicated that department leads experienced a number of common challenges. While in disciplines such as Psychology and Biosciences there was a strong existing understanding of the FAIR principles, in many other departments it was evident that the concepts had not yet become common currency. In such instances, department leads had a narrower range of existing good practice to draw from. Disciplines where data types tended towards the qualitative and sensitive often voiced concerns about how these data could be rendered FAIR without impacting either the need to protect vulnerable participants or the usefulness of that which was shared. Contrastingly, disciplines which often generated large quantitative datasets – Physical Geography, for example – raised questions about what amount and scale of data might be considered both manageable and useful to others. These points highlighted the degree to which one discipline may include a range of data types which each invite different forms and specifications of guidance that may best be addressed separately, a point we return to later. Further issues raised in relation to specific data categories included concerns about the compatibility of FAIR with commercially sensitive data (highlighting the need for permissions to be negotiated fully at an early stage) and relatedly, issues of data ownership where data originated in whole or part with a third party.

Other issues raised by colleagues in workshops and surveys included the lack of time and resources highlighted by Tenopir et al. (2011) and others and discussed above. This provided department leads and RDM colleagues with an opportunity to advocate for the importance of integrating planning for FAIR into the DMP process and, where the study is funded, making allowance in funding bids for the time and resources required to support FAIR processes. Additionally, colleagues across disciplines highlighted the need for further specificity on how to make their data and software interoperable. Some colleagues also expressed uncertainty regarding the degree to which FAIR practices were mandated at an institutional level, and in many instances were unaware of the support available to them in adopting the encouraged practices. This provided a valuable opportunity to highlight sources of support while also indicating that more work needed to be done at an institutional level and beyond in highlighting not only the FAIR principles but the mechanisms in place to support their uptake.

To turn to the resources that were developed, department leads were able to render these discipline-specific in a number of ways. The first related to the format of the resource: in discursive disciplines like English Language and Literature where, to risk caricaturing, researchers value fluent explication of ideas, the checklist took the form of a handbook in continuous prose, whereas in a pragmatic and process-oriented discipline such as Mechanical Engineering (with the same caveat), the checklist was presented as a series of short statements in relevant categories accompanied by tickboxes (see Wingham &Tipuric 2022). The community-led aspect of the resources' creation in this way enabled them to assume the forms most useful to

researchers in that discipline. Moving to the resources' content, departmental leads were able to focus the guidance on issues most pertinent to that discipline, with one example being the detailed guidance in the Mechanical Engineering checklist on exporting to a repository from GitHub. To a greater or lesser degree, as we discuss below, department leads also provided the necessary degree of discipline-specificity to make general suggestions such as repository use functionally applicable, as illustrated in the excerpt below (just part of the Biosciences checklist's guidance on discipline-specific repositories).

SUBJECT-SPECIFIC REPOSITORIES AND DATABASES

Genomics:

- **Gene Expression Omnibus (GEO):** https://www.ncbi.nlm.nih.gov/geo; data house for quantitative gene expression, gene regulation and epigenomic data, including data from RNA-seq, ChIP-seq, Hi-C, bisulfite sequencing and microarrays.

    - File formats: CRAM, BAM, SFF, HDF5, FASTQ, bedGraph, bigBed, WIG, bigWig, general feature format (GFF), gene transfer format (GTF) and GEOarchive

- **Sequence Read Archive (SRA):** https://www.ncbi.nlm.nih.gov/sra/; you can deposit high-throughput sequencing reads that do not fit into GEO.

    - File formats: CRAM, BAM, SFF, HDF5 and FASTQ

- **Genbank:** https://www.ncbi.nlm.nih.gov/genbank/: deposit DNA and RNA sequence data, which contains sequence of genomic DNA, mRNA, noncoding RNA, plasmids and synthetic constructs.

    - File format: FASTA

- **The European Genome-phenome Archive (EGA):** https://ega-archive.org/; deposits sensitive genetic and phenotypic information from human participants, allows controlled access to the data upon request.

    - File formats: CRAM, BAM, FASTQ, VCF, SFF and HDF5

(Zagrodzka & Simsek 2022)

Limitations in the creation of resources with this degree of specificity did emerge, however, in other disciplines where FAIR practices have less existing traction and there may be (for example) a lack of consensus within the discipline on appropriate places to deposit data, few or no discipline-specific repositories and a smaller pool of good practice to draw on.

EVALUATION AND CONCLUSIONS

This project resulted in resources of significant benefit to data management planning for FAIR in all of the pilot departments, albeit with some of the resources containing a greater degree of discipline-specificity than others. It also made significant strides in raising awareness both of the FAIR principles and of available support in data management planning and RDM at an institutional and sector-wide level. A number of questions emerged which will inform future activity in this area. First of all, it was evident that the FAIR principles were not sufficiently familiar to researchers in all disciplines for discipline-specific renderings of the guidance to be immediately useful – broad and introductory material was also required. The following question emerged: is a checklist the best format to accompany both general and specific levels of information in a concise and practical way, and what format of resource/s might best accommodate the needs of different audiences with different degrees of existing familiarity with FAIR? Questions also surround the feasibility of rolling out such a project at scale given the financial costs. Finally, a number of researchers consulted indicated a preference for centrally provided resources rather than those established within an individual community at the institution; this contradicts the expectation that community-led initiatives are most positively received by research communities, but would address concerns about the extent to which the resources would continue to be revisited and updated over time if these responsibilities were dispersed.

Bearing in mind these issues and concerns, a decision was reached to modify the approach at TUoS going forward in providing discipline-specific guidance to inform FAIR data management planning and RDM practices. Core resources will be developed on a platform such as Google Sites and will contain both general guidance on FAIR and specialised information by data type, which, as noted above, was in many cases a more relevant delineator than discipline. This resource will be developed to incorporate further specialised and discipline-specific guidance on FAIR via an iterative process of engagement with departments and research groups, initiated by Faculty-level workshops introducing the resource, encouraging peer discussion and advertising opportunities to work with a specialist to develop applications of these principles to specific departments or groups. The support and engagement offered to individual departments and/or research groups will include the following:

1.1. A talk at a department meeting and/or PGR training session.
1.2. A workshop on FAIR for departmental researchers to include exploration of discipline-specific issues. These will also be used to further develop resources, adding specificity and key examples of how researchers have applied FAIR principles within that domain.
1.3. Support to specific research groups and/or departments in customising the guidance for their own purposes (via a copy of the centralised guidance).
1.4. Support for the development of working groups/networks at a departmental level.

Alongside this resource and engagement, our aim is to further explore researchers' training needs via a range of methods alongside the institution's Open Research Training Lead; communicate clearly with researchers via multiple channels about available training and support; and apply the outputs of an ongoing UKRN project to develop training around open research (REF) once these emerge. As an RDM support team, we also aim to develop more detailed and specific training around data management planning and RDM for specific data types including qualitative and sensitive data.

This project illustrates some of the challenges in creating discipline-specific guidance to inform FAIR data management planning. It enabled insights into the obstacles encountered by researchers dealing with different types of data, as well as exploring how pragmatically such guidance might be developed and delivered at scale. Ultimately, it led to the development of future plans to roll out FAIR guidance at scale and may be of interest to other institutions seeking to develop and disseminate such guidance at an institutional level.

DATA ACCESSIBILITY STATEMENT

Data supporting this publication can be freely downloaded from the University of Sheffield Research Data Repository at https://doi.org/10.15131/shef.data.20496855, under the terms of the Creative Commons Attribution (CC BY) licence.

FUNDING INFORMATION

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Jenni Adams orcid.org/0000-0003-2420-0763
University of Sheffield, UK

Bev Jones orcid.org/0000-0002-3900-9374
University of Sheffield, UK

Helen Foster
University of Sheffield, UK

REFERENCES

Abele-Brehm, AE, et al. 2019. Attitudes toward open science and public data sharing. *Social Psychology*, 50(4): 252–260. DOI: https://doi.org/10.1027/1864-9335/a000384

Adams, J. 2022b. FAIR video case studies. The University of Sheffield. Collection. DOI: https://doi.org/10.15131/shef.data.c.6165894

Adams, J, et al. 2022a. Departmental FAIR checklists. The University of Sheffield. Workflow. DOI: https://doi.org/10.15131/shef.data.20496855

Calamai, S and **Frontini, F.** 2018. FAIR data principles and their application to speech and oral archives. *Journal of new music research*, 47(4): 339–354. DOI: https://doi.org/10.1080/09298215.2018.1473449

Chue Hong, N, et al. 2022. FAIR principles for research (FAIR4RS Principles v1.0). *Research Data Alliance*. DOI: https://doi.org/10.15497/RDA00068

Fecher, B, Friesike, S and **Hebing, M.** 2015. What drives academic data sharing? *PloS one*, 10(2): e0118053–e0118053. DOI: https://doi.org/10.1371/journal.pone.0118053

H2020 Programme. 2016. *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020 (Version 3.0)*. European Commission. Directorate-General for Research & Innovation.

Huang, X, et al. 2012. Willing or unwilling to share primary biodiversity data: Results and implications of an international survey. *Conservation letters*, 5(5): 399–406. DOI: https://doi.org/10.1111/j.1755-263X.2012.00259.x

Library Carpentry. 2019. *Top 10 FAIR data and software things*. Available at https://librarycarpentry.org/Top-10-FAIR/ [last accessed 24 November 2022].

McKiernan, EC, et al. 2016. Point of view: How open science helps researchers succeed. *eLife*, 5: e16800. DOI: https://doi.org/10.7554/eLife.16800

Munafò, M, et al. 2017. A manifesto for reproducible science. *Nature Human Behaviour*, 1(1): 0021–0021. DOI: https://doi.org/10.1038/s41562-016-0021

Norris, E and **O'Connor, DB.** 2019. Science as behaviour: Using a behaviour change approach to increase uptake of open science. *Psychology & Health*, 34(12): 1397–1406. DOI: https://doi.org/10.1080/08870446.2019.1679373

Perrier, L, Blondal, E and **MacDonald, H.** 2020. The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis. *PloS one*, 15(2): e0229182–e0229182. DOI: https://doi.org/10.1371/journal.pone.0229182

Pierce, HH, et al. 2019. Credit data generators for data reuse. *Nature*, 570(7759): 30–32. DOI: https://doi.org/10.1038/d41586-019-01715-4

Stewart, AJ, et al. 2021. Improving research quality: the view from the UK Reproducibility Network institutional leads for research improvement. *BMC research notes*, 14(1): 458–458. DOI: https://doi.org/10.1186/s13104-021-05883-3

Suber, P. 2016. Promoting Open Access in the Humanities. In *Knowledge unbound: Selected writings on open access, 2002–2011*. London, England: The MIT Press. pp. 331–340. DOI: https://doi.org/10.7551/mitpress/8479.003.0045

Tenopir, C, et al. 2011. Data sharing by scientists: Practices and perceptions. *PloS one*, 6(6): e21101–e21101. DOI: https://doi.org/10.1371/journal.pone.0021101

UKRN. 2022. Open research across disciplines: How the principles of open research can be applied to your discipline. Available at https://www.ukrn.org/disciplines/ [last accessed 23 November 2022].

University of Sheffield. 2021. *University of Sheffield Research Strategy Delivery Plan* (internal document).

University of Sheffield. n.d. Research software engineering (RSE). Available at https://www.sheffield.ac.uk/dcs/research/research-software-engineering [last accessed 23 November 2022].

Wilkinson, MD, et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3: 160018. DOI: https://doi.org/10.1038/sdata.2016.18

Wingham, J and **Tipuric, M.** 2022. '202207_MEC Fair checklist'. In Adams, J, et al. (eds.), 2022a *Departmental FAIR checklists*. The University of Sheffield. Workflow. DOI: https://doi.org/10.15131/shef.data.20496855

Zagrodzka, Z and **Simsek, D.** 2022. '202207_Biosciences Fair checklist'. In Adams, J, et al. (eds.), 2022a *Departmental FAIR checklists*. The University of Sheffield. Workflow. DOI: https://doi.org/10.15131/shef.data.20496855

Zhu, Y. 2020. Open-access policy and data-sharing practice in UK academia. *Journal of Information Science*, 46(1): 41–52. DOI: https://doi.org/10.1177/0165551518823174