

Three camps, one destination: the intersections of research data management, FAIR and Open

Open data, FAIR (findable, accessible, interoperable and reusable) and research data management (RDM) are three overlapping but distinct concepts, each emphasizing different aspects of handling and sharing research data. They have different strengths in terms of informing and influencing how research data is treated, and there is much scope for enrichment of data if they are applied collectively. This paper explores the boundaries of each concept and where they intersect and overlap. As well as providing greater definitional clarity, this will help researchers to manage and share their data, and those supporting researchers, such as librarians and data stewards, to understand how these concepts can best be used in an advocacy setting. FAIR and open both focus on data sharing, ensuring content is made available in ways that promote access and reuse. Data management by contrast is about the stewardship of data from the point of conception onwards. It makes no assumptions about access, but is essential if data are to be meaningful to others. The concepts of FAIR and open are more noble aspirations and are, this paper argues, a useful way to engage researchers and encourage good data practices from the outset.

Keywords

Research data management; RDM; open data, FAIR, open science



ROSIE HIGMAN

Research Data
Librarian
University of
Manchester



DANIEL BANGERT

Scientific Manager,
Knowledge Commons
Göttingen State and
University Library
University of Göttingen



SARAH JONES

Associate Director,
Digital Curation Centre
University of Glasgow

Introduction

The last 20 years have seen several shifts in emphasis and priorities in the area of research data management (RDM) and sharing. Research funder policies have developed and strengthened over the years from vague aspirations to enforceable requirements with compliance and monitoring activities. In particular there has been a shift in the rhetoric from focusing on RDM (notably through data management plans) to including data sharing and access. If such analyses were extended, recent policies would be shown to require data to be not only managed and open (where possible), but also FAIR (findable, accessible, interoperable and reusable).

2 Little in the research data field has gained such traction and universal acceptance as the FAIR data principles, conceived at the Lorentz conference in January 2014 and then consulted on and first published under FORCE11.¹ While interpretations of what it means to be FAIR or how FAIR an object is vary, nobody disagrees with what the principles assert. FAIR effectively packages ideas that have a long history in the OECD principles² and G8 Science Ministers statement,³ bringing these elements together in a concise and clear way, under an appealing acronym. It can open conversations with researchers and funders in ways that dull old data management never did.

RDM, FAIR and open are three overlapping but distinct concepts. Each brings a different emphasis and strength, and there is much scope for enrichment if they are applied collectively. RDM is the bedrock: if data have not been properly created and managed during the early stages of research, it will be very difficult to make them FAIR or open. The data ownership, documentation, formats and standards used will all affect the ability to share effectively, and these choices are often defined a long time before final outputs are made available.

'FAIR ... can open conversations with researchers and funders'

Data management enables FAIR and open sharing, while the principles of FAIR and open can act as inspiration to engage researchers in effective data management (see Figure 1). Researchers often want to be FAIR, and sometimes open; they are noble aspirations. Data management in contrast is akin to the ugly duckling – it is seen as menial grunt work that people know they *should* do but do not particularly want to engage in.⁴ By using the more appealing language of FAIR and open, we can engage people in data management too.

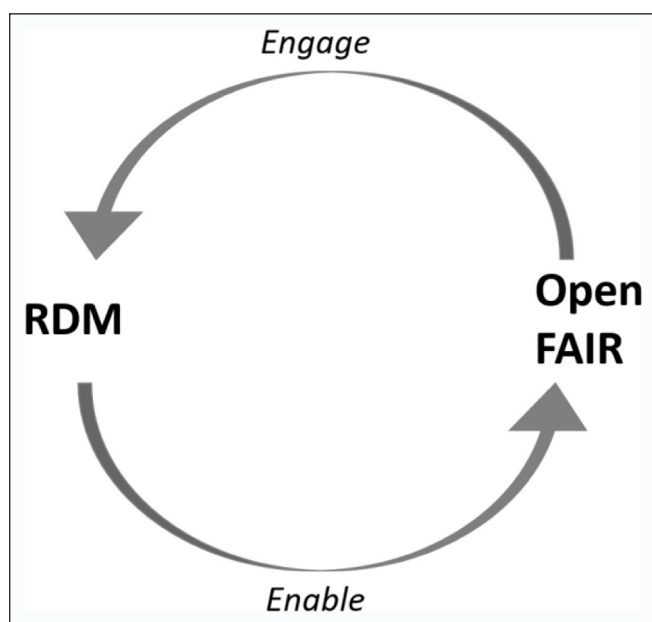


Figure 1. The virtuous circle of RDM, FAIR and open

RDM, FAIR and open

As concepts like FAIR are introduced, there is a need to address the relationship between it and other established ideas. Providing greater clarity around the intersections of RDM, FAIR and open can help to realize where alignment exists and identify gaps in awareness and support. This section will briefly review each of the three concepts and propose ways of understanding how they intersect.

Research data management (RDM) can be defined as a set of practices to handle information collected and created during research. It is 'the compilation of many small practices that make your data easier to find, easier to understand, less likely to be lost, and more likely to be usable during a project or ten years later'.⁵ These practices involve,

3 but are not limited to, data management planning, documentation, organization, storage, dissemination and preservation.⁶ Effective RDM is an ongoing process which is structured and aligned with the research context and disciplinary practices.

The FAIR principles advocate for increased findability, accessibility, interoperability and reusability of research data and scholarly digital objects more generally. Under the umbrella of the FAIR acronym, 15 principles have been formulated to guide the actions of data publishers, stewards and other stakeholders.⁷ Central to the concept of FAIR is its application 'to both human-driven and machine-driven activities', with a goal of machine-actionability to the highest degree possible or appropriate.⁸ In addition, FAIR is not binary (i.e. FAIR/unFAIR) but rather a spectrum along which varying degrees of 'FAIRness' are possible.^{9,10,11} While the FAIR principles have experienced swift uptake and acceptance, there are many directions to the current work connected to FAIR, including differing applications of FAIR to the assessment and implementation of services to support FAIR data.

'Central to the concept of FAIR is its application "to both human-driven and machine-driven activities"'

Open data is the practice of making underlying research data publicly available, accessible and reusable with minimal restrictions. Within the broader shift towards open science,¹² open data has increasingly become an expectation of funders and policymakers, often framed by the maxim of 'as open as possible, as closed as necessary'. Open data can be defined on a continuum, for instance by borrowing from Tim Berners-Lee's 5-stars of Linked Open Data (LOD).^{13,14} According to Berners-Lee, a minimum requirement of open data is to have an open licence (such as Creative Commons CC0), but to achieve greater openness and reuse potential, data should also be machine-readable, in a non-proprietary format, use open standards and link to other data to provide context. In this system, stars are accumulated by fulfilling each criteria step by step. These higher degrees of openness are where the overlaps with FAIR are most profound since both emphasize ways in which content can be made meaningful to support reuse by humans and machines.

Common misconceptions

There are a number of misconceptions about what RDM, FAIR and open mean. The terms are often conflated and used interchangeably. Here we try to unpack some of the most common misconceptions.

'One of the strengths of the FAIR principles is that they allow for controlled access'

FAIR data has to be open

No! While many policies call for FAIR and open data, the two do not mean the same thing. Data can be both FAIR and open, just one of these, or neither.

One of the strengths of the FAIR principles is that they allow for controlled access, which can be important for certain types of data. Both are also scales in which data or other outputs, such as code, can be made increasingly FAIR and open (see Figure 2).

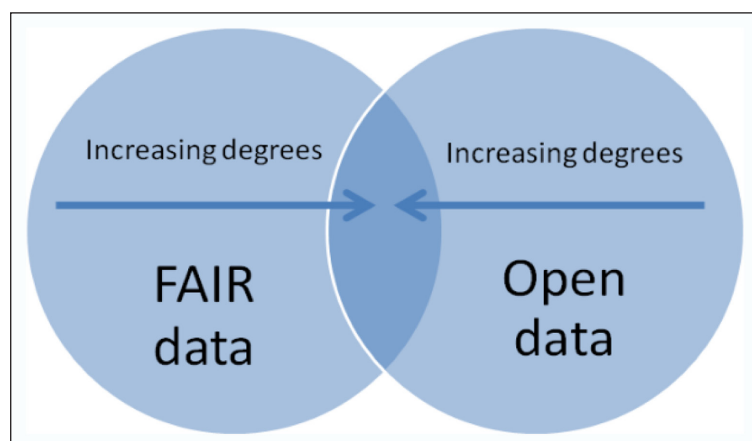


Figure 2. The relationship between FAIR and Open¹⁵

4 **Open data is more useful than FAIR data**

These concepts are not in competition; both are valuable and we should encourage researchers to make their data as FAIR and open as possible. The most reusable data will be well documented, conform to community standards and be as free from restrictions as possible to increase potential reuse.

All FAIR and open data is good quality

Neither FAIR nor open data are a reflection of data quality. Both are simply a measure of how data have been made available. A poor quality or fabricated data set could be both FAIR and open. This is why it is important to manage and document data well to provide the provenance and reassurances of how data have been created and processed, to engender trust. To be of most value, data should be well managed and provided with sufficient context to allow reusers to assess whether they meet their purposes.

FAIR is limited to the EU and the life sciences – why should I care?

Although FAIR grew out of a life sciences workshop in Leiden, the principles were intentionally articulated in a broad sense to apply to all types of data. Indeed, they are being applied in various contexts; the European Commission has put the FAIR principles at the heart of their research data pilot alongside open data.¹⁶ Beyond Europe, the American Geophysical Union (AGU) has a project on Enabling FAIR Data¹⁷ and the Australian Research Data Commons (ARDC) supports a FAIR programme.¹⁸

Modelling the relationship between RDM, FAIR and open

As outlined above, RDM, FAIR and open each have different emphases. Data management should not be subsumed by FAIR or Open as it deals with practices over a life cycle and has internal benefits to the researcher, project and institution which are not always related to data sharing as emphasized by FAIR and open. In particular, data quality issues are not covered by FAIR and open, yet are critical for reuse and supported appropriate data management and stewardship throughout the data lifecycle. RDM, FAIR and open are all important in their own right and should be viewed as complementary yet distinct.

'RDM, FAIR and open ... should be viewed as complementary yet distinct'

A way to conceptualize the relationship between RDM, FAIR and open is to consider each on a spectrum, as shown in Figure 3. This figure illustrates the intersections of managed, FAIR and open data in three-dimensional space.

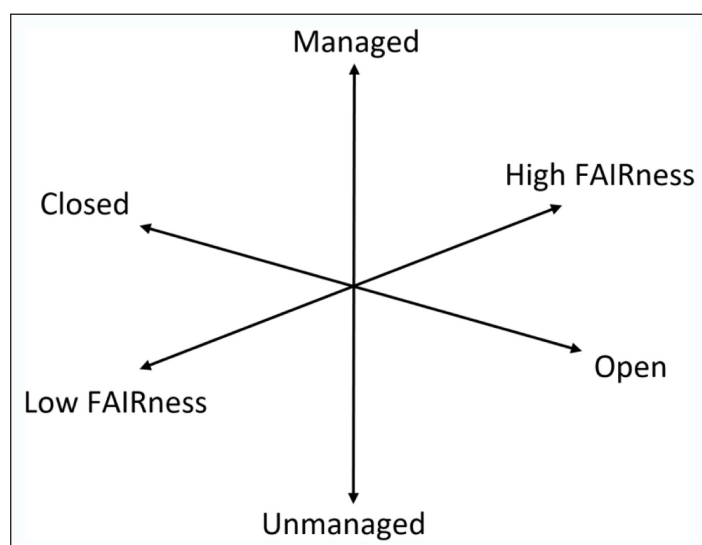


Figure 3. The relationship between managed, FAIR and open data

5 Our model of the relationship between managed, FAIR and open data recognizes variation along all three spectrums. In the model proposed in Figure 3, data can be:

- managed to varying degrees, from unmanaged to well managed
- open to varying degrees, from completely closed to highly open
- FAIR to varying degrees, from low to high FAIRness.

In general, the value of data are maximized when both openness and FAIRness are achieved to a high degree. Data rated as highly FAIR ought to have been well managed, but could be open or closed. In other instances, data could be made open or somewhat FAIR without being well managed, resulting in poorly documented and less reusable data. This is why it is important that data are also well managed to support sharing in a meaningful way and promote reuse.

Good data management is a necessary precursor for FAIR and open, and enables data to be created which is fit for sharing and reuse. Many decisions taken in the planning and management phases of research affect the potential for data to be made FAIR and/or open. These can include research project roles and responsibilities, consent agreements, data ownership and use agreed with partners, licences from third-party data owners, data format choices, metadata schema choices, naming conventions and the creation or capture of metadata and data documentation. By working from a foundation of effective RDM, researchers and data stewards can then consider what is an appropriate level of FAIRness and openness for the individual data set, taking into account factors such as content type, access condition, research project constraints and disciplinary practices.

Degrees of RDM, FAIR and open

To illustrate the intersections, boundaries and limitations of RDM, FAIR and open, two scenarios are discussed below. These demonstrate how these ideas can each support better stewardship of data in different settings and the respective limitations.

Partially FAIR and open but unusable data

One result of journal policies introducing data-sharing requirements is that more data sets are being shared. This does not always lead to reusable data, however. Open data sets may meet most of the requirements of FAIR whilst being practically unusable or of poor quality. A solitary CSV file with a limited description on a generalist data repository appears to tick lots of FAIR and open boxes (e.g. persistent identifiers, basic metadata, non-proprietary file formats, etc.) but limited documentation renders the data unusable without more information on provenance, explanation of the variables, and methodology.

Data may also be published as graphs and tables in image format or as supplementary files that cannot be directly manipulated and reused, such as PDFs. This does not mean that the creator has not managed the data well, rather that a reusable format has not been shared, often due to publisher policy. It is critical that we communicate the concepts of FAIR, open and RDM effectively so researchers understand potential limitations of supplementary files and recognize that data are a valuable research output in their own right. Data must be shared in editable formats and with sufficient documentation to allow them to be assessed, reused and potentially integrated with other data.

'Data must be shared in editable formats and with sufficient documentation'

Closed model data alongside FAIR and open code

In some disciplines, including engineering and computer science, the code and software being developed is frequently more important to the research than the data, which is primarily being used to test the code. In these disciplines, it is questionable to what degree the data should be managed and made openly available. This data, often termed model or synthetic data, may be unmanaged, closed and not adhere to the FAIR principles, whilst the

6 code can be highly managed, documented and made openly available. The flexibility in the FAIR principles means they are also easily applicable to code as it has many of the same properties as data including community standards, persistent identifiers and licensing.¹⁹ Thus, the FAIR principles can be used to have a helpful conversation around what is needed to improve the transparency and reproducibility of research, whether it primarily relies on data or code.

Advocating for RDM, FAIR and open

Managing and sharing research data are often not a high priority when talking to researchers, and whilst RDM, FAIR and open all help to encourage good practice, this proliferation of terminology can sometimes cause confusion. Careful thought is needed about how to use these concepts, and when. The suggestions presented below are a summary of major issues raised by practitioners at multiple events over the last couple of years, drawing heavily on a birds of a feather (BoF) session at the Engaging Researchers in Good Data Management Conference, Cambridge in 2017.²⁰ The discussion between librarians, data stewards and researchers at this event focused on how practitioners were using FAIR and open to advocate for effective data management. Five recommendations for using FAIR and open when advocating for RDM are summarized here:

1. Focus on FAIR when data cannot be open

One of the difficulties in advocating for open data has been that researchers with human participants, particularly those researching sensitive topics, frequently cannot share their research data, and may not be willing to ask participants for permission to share out of fear that it will deter people from participating in their research. FAIR can be extremely helpful here as it only requires that metadata are available. These could point to closed data sets, but ideally the data will be accessible in some form too. The fact that data do not necessarily have to be publicly available opens up discussions with researchers which were not previously possible. Participants in the BoF session mentioned using FAIR to encourage researchers to share a portion of their data or a metadata-only record which describes the data set in detail and outlines the steps for accessing it. This is particularly helpful following the introduction of GDPR (the General Data Protection Regulation), when researchers are arguably more concerned than ever about sharing personal data, and using FAIR to start this conversation allows for better RDM more generally to be promoted.

2. Acknowledge impact, citation and prestige

When discussing both FAIR and open data, many participants raised issues around citations, impact, prestige and researcher assessment. Although these issues may not be the focus of RDM, it is important to recognize the pressures on researchers to publish papers in a prestige journal, and be clear about how sharing data in a FAIR and open way can help support this. Here, the 'findable' in FAIR can help encourage good RDM by appealing to researchers' desire to make data more visible, thus incentivizing them to include sufficient metadata in a data set to make it findable. The emphasis on reusability in FAIR can also be linked to the impact agenda, encouraging data sets which are shared to be actively used in other contexts rather than just referred to.

3. Use terminology to help, not obscure or scare

Most researchers are already aware of 'open' as a concept via open access (OA), and this can help start conversations with researchers who have not considered sharing data before. This needs to be balanced against the association of OA with compliance and, in the UK, the Research Excellence Framework (REF). Open data is a relatively simple concept, at least on the surface, and one which can be quickly understood, if hard to implement in some cases. It is in the implementation of data sharing where FAIR can be useful as an advocacy tool. Whilst FAIR introduces another acronym requiring explanation, it can then be used to effectively walk researchers through the steps needed to make their data accessible and understandable. The 15 FAIR principles include several clear action points, such as obtaining a persistent identifier, assigning a usage licence and providing metadata online.

4. Consider the end-users of the data

One recurring theme is whether it was more important to maximize the amount of data available or to have fewer high quality data sets. When openness is the sole focus, there is a risk that lots of data sets shared may be of a low quality, with poor metadata, so that the data can only be comprehended by the researcher's immediate peers. By contrast, advocating for FAIR may result in fewer data sets being shared due to the increased requirements emanating from the principles, but they may be richer and more easily understood outside the discipline. Thus any conversation around sharing data needs to focus on the audience the researcher is aiming to reach and what possible uses there could be of their data, as well as making it FAIR and open. It is important to recognize the effort involved in managing and sharing data so that reasonable judgements can be made about when and where to apply this, as demonstrated in the model data example above.

5. Keep research and the researcher at the centre of the message

Whether using the terms RDM, FAIR or open, it is all too easy for those of us advocating for data management and sharing to get caught up in our concern for the data and forget what matters to the researcher – their research! We do not advocate for data management and sharing for their own sake, and whilst the end goals vary from reproducibility, to reuse across disciplines, to application by practitioners, the use which that data can be put to should be at the heart of our activities. So, whilst acronyms and familiar terms can be a helpful shorthand to describe the sometimes complicated data management practices we are encouraging, they should never replace a focus on improving the efficiency and inclusivity of the research process and helping new research questions be answered.

'the use which ... data can be put to should be at the heart of our activities'

Conclusions

RDM, FAIR and open are all important in their own right and should be viewed as complementary yet distinct concepts. All three exist on a spectrum and intersect with each other: data can be managed to varying degrees and be more or less FAIR and open. We should see each as a level of maturity in which researchers are encouraged to make their data more managed/FAIR/open, so it is ultimately more useful. Data management is the necessary precursor to enabling FAIR and open data and, conversely, these principles can help advocate for good data management practices.

Being FAIR and open is not necessarily sufficient. The internet was conceived as a mechanism for sharing content between trusted sites of authority. Anyone can be a data creator and publisher online. There are few controls to help users know which data can be trusted, hence the importance of professional curation in certified repositories to ensure data are effectively stewarded and remain accessible in the long term.

'Data management is the necessary precursor to enabling FAIR and open data'

The boundaries and intersections between RDM, FAIR and open cover important elements that risk being overlooked if we only focus on one concept. Properly stewarded FAIR data have much potential for reuse, but if they can also be made available as open data, this reuse potential grows. Similarly, if open data are uniquely identified so they can be discovered and professionally curated in the long term, the likelihood and depth of reuse will grow. We should advocate for data to be as FAIR and as open as possible, using these principles to help seed good data management practices from the start. The whole is greater than the sum of its parts.

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. FORCE11, "The FAIR Data Principles," <https://www.force11.org/group/fairgroup/fairprinciples> (accessed April 5, 2019).
2. OECD, *Principles and Guidelines for Access to Research Data from Public Funding*, 12 April 2007; DOI: <https://doi.org/10.1787/9789264034020-en-fr> (accessed April 5, 2019).
3. Foreign and Commonwealth Office, *G8 Science Ministers Statement*, 13 June 2013, <https://www.gov.uk/government/news/g8-science-ministers-statement> (accessed April 5, 2019).
4. Sarah Jones, "Open data, FAIR data and RDM: the ugly duckling," March 2018; DOI: <https://doi.org/10.1145/3158344> (accessed April 5, 2019).
5. Kristin Briney, *Data Management for Researchers* (Exeter, UK: Pelagic Publishing, 2015).
6. Louise Corti, Veerle Van den Eynden, Libby Bishop and Matthew Woollard, *Managing and Sharing Research Data* (London: SAGE, 2014).
7. FORCE11, "The FAIR Data Principles".
8. Mark D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data* 3 (2016): 3; DOI: <https://doi.org/10.1038/sdata.2016.18> (accessed April 5, 2019).
9. Barend Mons et al., "Cloudy, increasingly FAIR; Revisiting the FAIR Data guiding principles for the European Open Science Cloud," *Information Services & Use* 37, no. 1 (2017): 49–56; DOI: <https://doi.org/10.3233/ISU-170824> (accessed April 5, 2019).
10. Mark D. Wilkinson et al., "A design framework and exemplar metrics for FAIRness," *Scientific Data* 5 (2018); DOI: <https://doi.org/10.1038/sdata.2018.118> (accessed April 5, 2019).
11. "FAIR Data Maturity Model WG," *Research Data Alliance*, <https://www.rd-alliance.org/groups/fair-data-maturity-model-wg> (accessed April 5, 2019).
12. "Introduction: Open Science Training Handbook," *FOSTER*, <https://book.fosteropenscience.eu/en/01Introduction/> (accessed April 5, 2019).
13. "Design Issues: Linked Data," Tim Berners Lee, <https://www.w3.org/DesignIssues/LinkedData.html> (accessed April 5, 2019).
14. "5-star Open Data," James G. Kim and Michael Hausenblas, <https://5stardata.info/en/> (accessed April 5, 2019).
15. European Commission Expert Group on FAIR Data, *Turning FAIR into reality*, 26 November 2018, <https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1> (accessed April 5, 2019).
16. "Open Research Data the New Norm in H2020," *OpenAIRE*, <https://www.openaire.eu/open-research-data-the-new-norm-in-h2020> (accessed April 5, 2019).
17. "Enabling FAIR data project," Coalition for Publishing Data in the Earth and Space Sciences, <http://www.copdess.org/enabling-fair-data-project/> (accessed April 5, 2019).
18. "The FAIR data principles," *Australian National Data Service*, <https://www.ands.org.au/working-with-data/fairdata> (accessed April 5, 2019).
19. Neil Chue Hong and Daniel S. Katz, "FAIR Enough? Can we (already) benefit from applying the FAIR data principles to software? (Version 2)," December 2018, <https://doi.org/10.6084/m9.figshare.7449239.v2> (accessed April 5, 2019).
20. "Engaging Researchers in Good Data Management Conference 2017," University of Cambridge, <https://www.repository.cam.ac.uk/handle/1810/270234> (accessed April 5, 2019).

Article copyright: © 2019 Rosie Higman, Daniel Bangert and Sarah Jones. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use and distribution provided the original author and source are credited.



Corresponding author:

Rosie Higman

Research Services Librarian (Data)

Research Services, Red 1, University of Manchester Library

The University of Manchester

Oxford Road, Manchester M13 9PP, UK

E-mail: rosie.higman@manchester.ac.uk

ORCID ID: <https://orcid.org/0000-0001-5329-7168>

Co-authors:

Daniel Bangert

ORCID ID: <https://orcid.org/0000-0003-4981-2870>

Sarah Jones

ORCID ID: <https://orcid.org/0000-0002-5094-7126>

To cite this article:

Higman R, Bangert D and Jones S, "Three camps, one destination: the intersections of research data management, FAIR and Open", *Insights*, 2019, 32: 18, 1–9; DOI: <https://doi.org/10.1629/uksg.468>

Submitted on 08 March 2019

Accepted on 01 April 2019

Published on 22 May 2019

Published by UKSG in association with Ubiquity Press.