

# IMPLEMENTATION

# STORIES



**FAIRSFAR**  
Fostering Fair Data Practices in Europe



THEME  
**2**

**DEVELOPING SUSTAINABLE  
BUSINESS MODELS**

# IMPLEMENTATION STORIES

THEME **E** DEVELOPING SUSTAINABLE  
BUSINESS MODELS

## Evolving research lifecycle support through data stewardship and research software engineering

### Authors

Angus Whyte, Ryan O'Connor (Digital Curation Centre) and Vasso Kalaitzi (DANS)

### From interviews with

Bill Ayres, Strategic Lead, Research Data Management, University of Manchester, Dr Danielle Owen, Research Data Manager (Core Facilities), Faculty of Biology, Medicine and Health, University of Manchester, and Dr Peter Crowther, Systems Administrator, Research IT, IT Services, University of Manchester

MANCHESTER  
1824

The University of Manchester

## Introduction

This Implementation Story looks at research data management service provision at the University of Manchester, and how their service offering has developed over time in response to researcher needs, changes in the RDM environment, and the drive to help researchers make their data more FAIR. Based on an interview with three University of Manchester staff involved in RDM at the institution, the Story charts how capabilities for data stewardship and research software engineering are increasingly coordinated with a mature RDM service, detailing the strategic goals, roles, workflows and further service implementations.

This story aligns with the FAIRsFAIR D3.4 recommendations on practice to support FAIR data principles:<sup>1</sup>

### **FAIRsFAIR recommendation**

*"Develop and implement models for coordinating and supporting data stewards and research software engineers."*

**FAIRsFAIR Recommendations on practice to support FAIR principles**

1. Molloy, L., Nordling, J., Grootveld, M., van Horik, R., Whyte, A., Davidson, J., Herterich, P., Martin, I., Méndez, E., Principe, P., Vieira, A., Asmi, A. (2020). D3.4 Recommendations on practice to support FAIR data principles (1.1). Zenodo. <https://doi.org/10.5281/zenodo.5357329>

## Research Data Management at the University of Manchester

The University of Manchester has been evolving its support for research data management (RDM) over more than 10 years. The work around RDM service provision initially focused on providing faculty-level support for researchers, initially to those working in life sciences domains, for example, in biomedical sciences,<sup>2</sup> before expanding to other areas of the university. In this time, RDM services have matured, with the university library one of the driving forces behind this. With a broader focus on data stewardship, the University is now looking to refresh and renew its coordination of support across faculty-level support facilities, beginning where it started - in the life sciences.

From its central position in the university library, the RDM service provides researchers with access to customised data management plan (DMP) templates via its institutional subscription to the DMPonline platform<sup>3</sup> and its own institutional data repository via figshare<sup>4</sup>, in addition to support in the form of training and general advice on all RDM matters, including directing researchers to external sources of discipline-specific support, e.g., ELIXIR for life sciences<sup>5</sup>. However, as Bill Ayres, Strategic Lead for Research Data Management in the university library, says, these developments have not happened in isolation: "We are advocates and help with policy and advice, but we also work very closely with our colleagues in Research IT and IT infrastructure teams, and indeed with colleagues in faculties, especially with experimental facilities operating at the coalface of data production and management in the university."

More recently, the university has undertaken a plan to significantly uplift all aspects of research support, taking into account researchers' needs and removing the barriers they face in carrying out their work. The Research Lifecycle Programme (RLP) is investing £15 million over five years, with a focus on understanding and helping to improve research processes, as well as developing the necessary research support systems and e-infrastructures.<sup>6</sup>

Early scoping for the RLP identified the need for an overarching strategic vision for RDM at the university. It was in response to this that Bill's role was created, working within the RLP but also within the library service to steer the various RLP projects, towards tangible outcomes for the RDM service, platforms and infrastructure.

One of the elements of the current RDM service provision that the RLP is looking to address is to renew and coordinate the faculty-level focus that characterised the initial stages of RDM at the university. Across the different faculties there are many people in different roles working with researchers in different ways – those in researcher-facing roles, in supporting services, or in data specialist roles at faculty level. "Historically," Bill notes, "they're probably more separated than they would like to be." Part of the strategic vision for RDM at the university involves bringing these under the umbrella of 'data stewardship'. This aims to help those in such roles to understand what data stewardship involves, in turn helping to foster communication between those in similar roles in different domains and ensuring that the university's central RDM service is in touch with the disciplinary-specific groups.

Bill cites the work being done by the Research Data Alliance Professionalising Data Stewardship Interest Group as useful in this respect, providing models for institutional data stewardship and the relevant terminology for clear role definition, which will potentially unlock the adoption of data stewardship at a high level across the university and aid in getting buy-in from senior management.<sup>7</sup>

2. Poschen, M., Finch, J., Procter, R., Goff, M., McDerby, M., Collins, S., Besson, J., Beard, L., & Grahame, T. (2012). Development of a pilot data management infrastructure for biomedical researchers at University of Manchester – approach, findings, challenges and outlook of the MaDAM project. *International Journal of Digital Curation*, 7(2), 110–122. <https://doi.org/10.2218/ijdc.v7i2.234>

3. <https://dmponline.dcc.ac.uk/>

4. <https://figshare.manchester.ac.uk/>

5. <https://elixir-europe.org/>

6. <https://www.rlp.manchester.ac.uk/>

7. <https://www.rd-alliance.org/groups/professionalising-data-stewardship-ig>

## Developing the workflow and roles

### Scaling up faculty-level services

---

There are nine core facilities within the Faculty of Biology, Medicine and Health at the University of Manchester, and these collectively produce very large amounts of data, roughly a petabyte of data annually. Hence a decision was made to address RDM at source by trying to implement best practises within the core facility workflows. As mentioned above, up until recently each core facility has been responsible for handling their own data and workflows, which led to very different approaches to RDM, which is unhelpful for managing risks regarding data storage and findability.

This situation is being addressed by setting up an integrated data storage and compute platform, to which all core facilities can send their data. This data can then be accessed by researchers via hosted virtual machines, which are already pre-loaded with the software needed for the data analysis. Crucially, the raw data is protected and held by the core facilities. The researchers can access the raw data to perform their analyses, and then store the latter in their personal PI (Principal Investigator) data storage. This approach aims to reduce the risk of raw data being stored on personal devices and hard drives, in Dropbox folders and shared via email, this process has the key advantage of making data findable to those within the research group, as well as core facilities staff, who can access and retrieve data should a researcher leave their position.

Danielle Owen, whose time is divided between her roles as research data manager at the faculty of biology, medicine and health and as operational lead for the university's data safe haven service, explains that the team at University of Manchester also recognised that there was an opportunity to make the data more interoperable. Having centralised data storage and computational functions made this more straightforward, taking the data produced in proprietary formats and manipulating it: "We can convert that data into standardised, open formats on the fly; we have a script that runs in the background to do that for us. There are only some instruments that are compatible with that, others are not, but where possible we're trying to do that."

Furthermore, the team works towards capturing metadata as well. There are two aspects of metadata to take into account: the metadata that the core facilities capture and the metadata collected from researchers. To collate those two sources of metadata, the team are investigating the use of the Open Science Framework (OSF) platform.<sup>8</sup> Another initiative involves using Microsoft Word forms with drop-down lists for standard metadata elements to avoid people having to fill in information repeatedly. This approach comes with a series of videos explaining to researchers how to create their own forms which link to the relevant data folders and capture information to help make the contents reusable.

### Research Software Engineers Group

---

The RDM team identifies 'data stewardship' with a variety of roles occupied by people who are not 'data stewards' as such. The overall aspiration is to have a formalised approach to data stewardship at the University of Manchester, to be able to link up initiatives and avoid duplication of effort and formalise the support roles around most commonly requested services.

In contrast, a more formal recognition of the Research Software Engineers (RSEs) role has already been achieved at the university. Similar to those in other data stewardship positions, RSE roles have developed organically in the last decade in response to researcher needs. An RSE group has been in existence at the university for the past five years. Peter Crowther was active in this group and sees benefits in having this more formal definition.

8. <https://www.cos.io/products/osf>

# IMPLEMENTATION STORIES

THE IAC DEVELOPING SUSTAINABLE  
BUSINESS MODELS

Funding bodies are beginning to give the role proper recognition, and this has assisted in proper budgeting and costing in grant proposals: “Essentially, this means that all of the RSEs in the group are supported by research funding. We receive a small amount of central funding from central IT services, but for the most part, we are cost neutral, which is great, because this means we are less susceptible to the periodic reallocation of central university funding.” Though again, similar to those in data stewardship roles, there is still some way to go to align the service with researcher needs, as Peter notes: “We’re still working on formally defining our roles and responsibilities. At the moment, our service is often just whatever the customers want us to do.”

## Transferable methods for solving visible problems

### ■ Engaging SHAPE researchers

---

Some aspects of research software engineering that Peter Crowther and colleagues have found helpful in establishing a more formalised approach are less evident for data management support roles. These include the transferability of the skills to apply software engineering methods, researchers’ self-awareness of the gaps in their own skills, and the visibility of the problems that software addresses. For example, researchers are aware of issues with data storage and security, and these can be addressed through transferable software approaches. While data practises may vary widely across disciplines, Peter Crowther notes that software requests do not vary as much as might be expected, for example, comparing physics and astronomy to the humanities. What matters more is how sure the researchers are of what they need: “From what I have seen, it varies according to the individual, not necessarily the faculty.”

From a broader perspective, making the case for the benefits of research software engineering is comparatively more straightforward than communicating the benefits of good RDM. An advantage of software, according to Peter, is that it is more visible: you write code, and the code produces results. Poor data management can persist for longer periods without it being noticed, which makes the issues less visible, although both good data stewardship and software engineering can facilitate higher quality research.

The timing of support requests for both RSEs and data stewards is affected by these issues. Danielle says that, while some researchers do seek out support proactively, there is still a lot of ‘firefighting’ that goes on by those providing RDM support: “We are pushing now to engage the core facilities when we become aware that a research PI is submitting a grant.... The facilities are generally quite happy to have that because they also like the PIs to be able to write about the data management practises that we’re putting in place in their DMPs.”

Getting buy-in from researchers starting out in their careers makes a big difference. This has motivated University of Manchester to consider their training offering and how it can be more effective, for example by including a broader perspective of data stewardship and FAIR. Connecting with cohorts earlier in their careers (e.g., early career or postgraduate/postdoctoral researchers) will make it more likely that RDM and FAIR skills will be ‘hardwired’ into the work they do. As Bill notes: “For people that are starting out in their careers, [this] makes a big difference in how they work with data, how they manage data as they go on. We’re currently looking at our training offer and thinking about how we can make this work more effectively, and obviously includes things like FAIRness.”

## Supporting FAIR through a Research Data Repository and other

The FAIR principles have explicitly motivated initiatives of the central RDM service as it is delivered through the research lifecycle programme. As mentioned above, the university renewed its data repository, launching a Figshare-based service in 2021. The team are hoping to encourage more researchers to make use of this facility over the next few years, with the focus on making datasets more Findable and Reusable.

A further development the team expects to help enable FAIR is a plug-in to link the repository with the university's current research information system (CRIS), which is based on Pure. This integration will present datasets from figshare in the researcher's Research Explorer profiles. It will also interrogate and bring datasets in from most external repositories. Bill explains that "the emphasis here is on taking the wide spread of data deposited from Manchester researchers in disciplinary and general repositories all over the world, and then being able to present that in their research profile, so that it brings much better visibility for those data resources.... Once we make that move it's going to really make a big difference to how we think about FAIRness for Manchester data."

## Supporting cost budgeting

As with all models of RDM service provision, the recovery of costs needs to be considered. Like many institutions, Manchester has sought to simplify cost management from the researcher and research group perspective. The university provides projects with a set amount of 'free' storage, which in most cases is enough for their needs. Where projects are known to be particularly data-intensive (Danielle gives electron microscopy as one example) PIs need to be approached to make sure that they are costing the extra data storage they need into their grant proposals. There is also a mechanism in place using the university's DMPonline platform subscription. When researchers answer specific conditional questions set up on the platform, this sends out requests for additional storage where necessary and a notification that this additional storage needs to be costed into any grant proposal.

When it comes to staff costs for RDM support there are slightly different approaches. Danielle's salary cost is distributed across core facility orders resulting in a small increase in core facility charges, which goes unnoticed. Research software engineer support is similar to data support roles, in that researchers or projects can get a certain set amount of free support, usually a day or two. After that, support is still available, but it has to be costed into the project's budget, with the group charging a flat day-rate fee for any of its specialists assigned to a project.

The key thing for the RSE group, according to Peter, is that originally the group was supported by IT central services, with permanent positions funded. This initial 'bootstrapping' phase allowed the group to build up capacity and capability through the retention of experienced staff and the ability to build customer relations in the longer term, both of which result in being able to offer a better service. The budget for RSEs' time can be allocated to individual research projects, while the RSE positions are independent of any specific projects, with these permanent positions still sitting within IT central services and not tied to any particular grants.

# IMPLEMENTATION STORIES

THEIR DEVELOPING SUSTAINABLE  
BUSINESS MODELS

## Future focus

Much of the work outlined above – encouraging use of the institutional repository, fostering better RDM practises, defining a cost model for support - are ongoing activities. A final issue that was discussed in the interview focused more on the long term. How to go about archiving and preserving datasets for 10-plus years is now a consideration. Bill draws the distinction between datasets that explicitly support another research output (e.g., paper publication), and other datasets not linked to any other research outputs. Although the former has been the priority for depositing with the university repository there is a broader category of data that needs to be retained, archived and preserved beyond the end of a project. At the moment, an institutional-level preservation option does not exist, but looking to the future, Bill notes this is “going to be a big focus for the next couple of years from a central RDM service perspective.”

## ■ *About FAIRsFAIR Implementation Stories*

---

FAIRsFAIR Implementation stories illustrate good practices in research communities and organisations to support the implementation of the FAIR principles. These practices encompass 'FAIR-enabling' actions as recommended in the EC Expert Group on FAIR report Turning FAIR into Reality and the FAIRsFAIR Recommendations on practice to support FAIR principles. FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

## ■ *FAIRsFAIR - Fostering FAIR Data Practices in Europe*

---

European Commission Grant Agreement No. 831558  
H2020-INFRAEOSC-2018-4  
[www.fairsfair.eu](http://www.fairsfair.eu) - [support@fairsfair.eu](mailto:support@fairsfair.eu)

## ■ *Acknowledgements*

---

This report has been produced by the FAIRsFAIR (GA No. 831558) project, which received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-05-2018-2019. FAIRsFAIR gratefully acknowledges everyone who we could interview for this series.

## ■ *Disclaimer*

---

The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content. **January 2022**

## ■ *Copyright rests with the authors*

---



This work is released under a Creative Commons Attribution License, version 4.0.  
For details please see <https://creativecommons.org/licenses/by/4.0/>

**JOIN OUR COMMUNITY!**

 @FAIRsFAIR\_eu  /company/fairsfair

[www.fairsfair.eu](http://www.fairsfair.eu)



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.