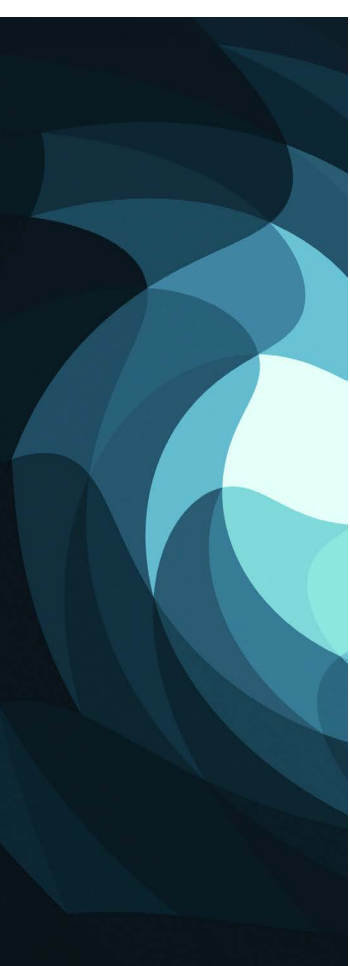


# Common themes in data management emerge at the MRC's open data workshop

A FAIRy tale by [Phillip Crout](#) – a personal story about how I work with FAIR data



In October of 2022, I hosted an all-day seminar at the Medical Research Council Biostatistics (MRC) unit to give a community of like-minded researchers a chance to discuss and explain the implications of open-data practice for their research<sup>1</sup>. The day was a resounding success, with a number of excellent talks and conversations. One thing I found particularly interesting – as a member of the research support team – was how often the same themes re-emerged despite the range of research interests represented. This feeling has only strengthened during my time as a fellow.

While all the conference attendees were from fields broadly considered ‘the life sciences’, the event had an interdisciplinary feel. Even so, throughout various applications, the data needs (and challenges) remained remarkably consistent. Many attendees had heard of the FAIR principles<sup>2,3,4</sup> before, but those that were new to the concepts were surprised by how well FAIR captures the challenges they faced. The dominant feeling over our four sessions was that often, re-

searchers get half/most of the way with their open data effort but then undervalue how important those final steps are.

One ‘final step’ that was often missed was providing long-term hosting solutions for data, and several conversations revolved around how important this is. We heard several horror stories of important data sets that were (in theory) ‘available on request from the corresponding author’ but in actuality, could not be accessed, sometimes as close to 6 months after publication date. Similarly, several attendees had navigated to self-hosted websites that gave 404 errors within a year of being mentioned in published research. While this might seem disheartening, this is something of a solved problem with Figshare<sup>5</sup> and Zenodo<sup>6</sup> providing hosting platforms that will last for as good as forever.

The presence (or absence) of appropriate metadata was a second key finding of our workshop. This was particularly important to researchers using genetics data as minor convention/naming differences could cause several hours to be lost when

trying to run the data through a new algorithm. Unfortunately, it can be difficult to establish one-size-fits-all metadata principles as the required information tends to be very field-specific. In an ideal world, everyone would provide 'as much metadata as possible', but several colleagues were keen to point out that this is often impractical and leads to crude full data dumps in which useful information is drowned out by unhelpfully large numbers of files containing information of little to no value.

The workshop had a huge range of perspectives, and I found the day fascinating. I hope this FAIRy Tale has helped to illustrate that doing the basics really well can be a huge source of value for both you and future researchers. In particular, it's essential to use a long-term, stable hosting platform for as much of your research data as possible and to provide accurate, complete and (ideally) concise metadata so that a new researcher can get up to speed with your data as quickly as possible.

## References

- 1 Making your research data open. UKRI. <https://ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-data-open/>
- 2 Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
- 3 Carpentries course: FAIR Pointers. <https://elixir-uk-dash.github.io/FAIR-Pointers/>
- 4 Carpentries course: FAIR in (biological) practice <https://carpentries-incubator.github.io/fair-bio-practice/>
- 5 Figshare. <https://figshare.com/>
- 6 Zenodo. <https://zenodo.org/>



### Phillip Crout

An ELIXIR-UK Fellow from the 2nd Cohort (2022-2023).

Phillip is interested in breaking down barriers between researchers and their open science goals. His journey began with academic software but has moved to more holistic approaches, hosting seminars and creating resources on data management. In addition, he is a culture leader within his institution, expressing his passion for a more just scientific community whenever possible.



This work was funded by the ELIXIR-UK: FAIR Data Stewardship training UKRI award (MR/V038966/1) 6/1)