Achieving FAIR Data With Figshare

[1]Dr Mark Hahnel
Figshare, London, UK[1]
mark@figshare.com

The FAIR (Findable, Accessible, Interoperable, Reusable) data principles are being adopted by funders, governments, publishers and institutions globally. But what are they? Are they a realistic expectation? How can researchers ensure their research data are FAIR?

The term "FAIR" was launched at a workshop in The Netherlands in 2014 and in 2016, the "FAIR Guiding Principles for scientific data management and stewardship" were published in Scientific Data [1]. The authors intended to provide guidelines to improve the findability, accessibility, interoperability, and reuse of digital assets for both humans and machines. The FAIR principles are a noble aspiration and could one day become the gold standard of data publishing but they are a far stretch from the reality of data publishing today.

The achievement of FAIR data can be grouped into 2 categories:
1 - Where the data is deposited
2 - How it is described.

As a platform, Figshare has came out on top of many independent studies of repositories and their compliance with the FAIR principles [2]. The second part of FAIR compliance pertains to the metadata associated with data and this is where things get very complex, very fast. Due to the vast array of digital assets being made available by researchers across the diverse scientific landscape there are very few standards governing what is acceptable metadata.

A further stipulation of the principles is that all data should be FAIR for both humans and machines for automated, computational analysis. "The principles emphasise machine-actionability (i.e. the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data." [3]

Current technology and the fluid status of metadata standards do not yet allow for machines to properly interpret the interoperability and reusability of data. That's where we believe there is an important role to be played by data curators within universities and publishers. Whilst platforms can do the heavy lifting regarding persistent identifiers, search indexing, knowledge representation and providing appropriate schemas for metadata, ultimately it is down to the subject experts to ensure the data is described in a way that can be interpreted by fellow researchers.

This presentation will walk through each of the FAIR principles and explain how by depositing data on Figshare, it ticks many of the FAIR boxes. We'll discuss the gaps that currently exist in the data publishing environment and touch on the critical next steps we need to take as a community in order to make FAIR a reality.

References

1 - The FAIR Guiding Principles for scientific data management and stewardship
https://www.nature.com/articles/sdata201618

2 - Evaluation of data repositories based on the FAIR Principles for IDCC 2017 practice paper
https://data.4tu.nl/repository/uuid:5146dd06-98e4-426c-9ae5-dc8fa65c549f

3 – GO FAIR – FAIR principles https://www.go-fair.org/fair-principles/