

## FAIR Principles

Keith Russell, Australian Research Data Commons (ARDC)

### Filmed at the Taylor & Francis Melbourne Scholarly Summit, 12 February 2019

First of all, who's heard of the FAIR data principles? Okay, so this was back 2014, Lorentz workshop in Leiden in the Netherlands. A group of thought leaders in this space, including publishers, including researchers, including scholarly societies, came together and said, "well, what is a good way of thinking about data and making data available?". And they came up with these, the four principles, making data findable, accessible, interoperable, and reusable. And they actually broke those down quite thoroughly. And it's proven to be really useful as a way of thinking about data. And I think one of the big extra bits of value there is that they're thinking about how data can be made available in a form that it can be also be picked up by machines and be brought together, found and reused.

So first of all, data should be findable. And when they say findable it means it should have a DOI over the data, it should be in a repository so that it can be found through different finding mechanisms around the world globally, so through machines. So preferably in a discipline repository or an institutional repository that feeds up into international collection mechanisms.

The data should be accessible. Now this doesn't mean that it has to be open. So there are good reasons why some data sets can't be made open, and that's especially around sensitive data. So data about humans, but also data about biological specimens. For example, the location of the Wollemi Pine. Well, now the Wollemi Pine is no longer threatened, but when it originally was found. That's the sort of data you don't want to throw out in the public immediately. You need to make it available through appropriate mechanisms so that the researchers that want to use it and want to be able to untangle and use those data, can get access to it, but that it is available through appropriate routes. The data should also be accessible in a format that's easy to pick up, sort of standard protocols that machines can use to find it and access it.

The data should be interoperable, as in using as much as possible discipline standards, so that other researchers can easily pick it up and don't need to find some proprietary software to read it, or use it, or be able to combine it with a different type of data.

And finally the data should be reusable. And when they talk about reusable, first of all, I always say the F, the A and the I are part of it being reusable. Because if you can't find it you won't be able to reuse it. But building on top of those points I mentioned earlier, it's important that the data also has a clear license over it, so that somebody knows what they can do with it. If there's no license, you don't know what you can do with the data. And also it needs associated information about how it was created. So just having a table there saying these are all our observations, you actually need to have a bit of context around that, knowing under which conditions was the data collected. Or if you have a series of interviews, under which conditions were those interviews conducted and why did you select those respondents for those interviews. So having some provenance information, how did you make that selection, and how did you select that data and make changes to the data maybe, is a very important aspect in being able to reuse that data and understand it.

One of the actions was we worked on at the AGU Enabling FAIR Data Project, we were one of the partners in that space, and that resulted in a commitment statement that has now been signed by a

range of repository managers, researchers, research institutions, and publishers, that sort of have committed themselves to making data available. And what they argue for there, and that's something we very strongly encourage, not to have the data as an appendix in the article, because that actually doesn't make it very reusable, but to have it in a repository where it can be made findable, accessible, interoperable, and reusable. So that means that other people can easily find that data. But having an appropriate DOI over that data with an appropriate citation, so that from the article you can reference the data with the appropriate credit associated with that.

When we're talking about making data accessible, you can talk about it on the short term, but having it in that repository also means that it can stay there for the longer term. And that's where we've been working with a number of data repositories and custodians around Australia, saying, "well, what does it mean to make that data available for the longer term and accessible for the longer term?". And one mechanism for that is to have it in a certified, trusted data repository, and there is now Core Trust Seal, an international standard certification organization, and you can have your repository certified as a CoreTrustSeal certified repository. So we've been working with the Australian Data Archive, great archive in Canberra, for social science data, and worked with them, what does it mean to have their repository certified as a CoreTrustSeal certified repository, and they were successful.

We've also been trying to provide guidance to the sector around Australia. What does it mean to make your data FAIR? What sort of things can you do? This is a self-assessment tool that researchers or research support staff can use to hold up a data set and see how well are we scoring on making this data FAIR, and what could we do to make it more FAIR? What steps could we take? What actions could we do? So it's just tangible, trying to turn it into real actions.

Also developing skills around the country, especially amongst the research support staff. And one of the things we did there was did a global sprint to develop top 10 FAIR data and software things in different disciplines, and they've been developed and they're now publicly available so anybody can pick them up and use them to untangle "what does it mean to make my data FAIR" in a specific area.

So to make data FAIR actually requires actions from a number of different players, stakeholders in this space. So what can a researcher do? A researcher can make sure that they have an ORCID, that they know their institutional repository or discipline repository. Make sure that when they publish their data it does have a DOI so they can reference it and think about that longer term. Editors can promote the use of trusted repositories. The actions you're already doing around standardizing journal data policy just makes it much easier for all the different stakeholders to publish that data in a consistent way. And if you look at the research supports staff, that's where they can help researchers make it easy, make it reasonably seamless, not to have to put too much effort into this space.