figshare • Particle Physics

# The value of a DOI and an open platform for sharing data

Tom Whyntie, Queen Mary University of London

I am a particle physicist. I did my PhD on the Large Hadron Collider's search for dark matter, which we didn't find, but we still haven't, so that's fine.

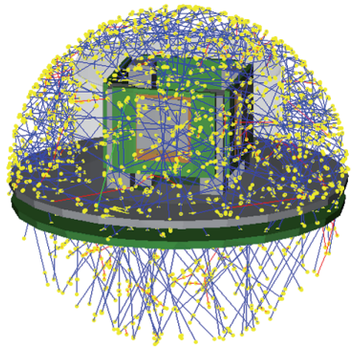
I'm now the Public Engagement Fellow for The Science and Technology Facilities Council and one of the major remits of that, which also ties in with some of the work that I do with the (Tim) Peake collaboration which does all the computing for the Large Hadron Collider, is about getting people engaged in actually doing research rather than just hearing about it or having it told to people online.

*" The key thing for me is the DOIs (digital object identifiers) which allow you to then cite that data in publications. "*

One of the major focuses for the work that we do is something called the Institute for Research in Schools, which is a charitable trust that offers school students and teachers the opportunity to do research. As you can imagine, a big barrier to that is access to facilities and equipment to start with, and even if you get past those things, you have issues with data access, data storage, and data management. And so, the reason I personally find figshare so exciting is that it does offer this open platform for sharing data and for collaborating with others without barriers.

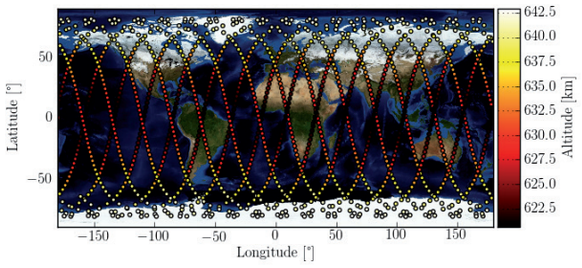
The key thing for me is the DOIs (digital object identifiers) which allow you to then cite that data in publications, which is the ultimate goal of the Institute for Research in Schools - to get students publishing peer reviewed content. So personally, I see figshare and repositories like that as one of the key tools that our students, and indeed anyone, can use to collaborate and do better science.

I'm primarily uploading datasets that aren't the subject of a current paper or publication because of the embargo period. It's really easy to upload and really easy to share. What's really nice is having the DOI that makes it citable, but also the URL that you then share. It makes it look official - it's not like some it's just a Dropbox link. That's actually quite important when you're working with schools or other research groups that it's a legitimate repository that you can then annotate with whatever supplementary information you need and you can upload .zip files. Most of our data from the detectors is in ASCII text format, which is quite nice because it's easy to read, but you can just zip it all up, pop it onto figshare, and it's there. People can download it and do what they want with it.



Allpix simulation of the LUCID experiment with 100 source protons

The other thing that's useful is, as the name suggests, figures. All of our publications have to be Open Access and one thing I'm particularly keen on getting right is copyright and making sure things are released under the Creative Commons license, because the key thing for schools and students is being able to reuse images without having to worry about copyright or commercial issues. So when you upload something, for instance one of the high quality photos of the detector, the arrangement, and the apparatus that I've uploaded, can be used under the terms of that CC license, which removes a whole level of headache.




Graphical representation of LUCID's orbit

So having that all built into the system is really nice. The same goes for the teacher resources for things like Timepix or an experiment guide for the eclipse experiment we did last year - again, not only is it there for people to download, but the download tracking feature is nice, as well, so we can monitor how often they're used and cited.

*" If you know you've got somewhere you can upload [your research], with a DOI in a repository that conforms with the Open Access policies of your institution, that then becomes part of your data management plan."*

figshare has changed the way I share my research in the sense that if you know you've got somewhere you can upload it, with a DOI in a repository that conforms with the Open Access policies of your institution, that then becomes part of your data management plan. It crystallizes that and you have a very clear idea of where that will go. It makes you think more about your data and how it is documented and presented. Often you think, "Oh, actually, I haven't done that or I haven't thought about how that does link up with that", which generally makes the dataset better and can also help the analysis as well because if there are bits that aren't clear, it can help avoid errors because you're making sure it's good enough to put out there. So it can actually improve the quality of the science, as well. To have another point of doing that other than the publication of the paper can only be a healthy thing.