Figshare: Interview with Mark Hahnel

Martin Fenner, Gobbledygook

February 16, 2012

fig**share** *allows researchers to publish all of their research outputs in seconds in an easily citable, sharable and discoverable manner.* The service was started by Mark Hahnel last year while still a PhD student. Mark joined Digital Science to work on fig**share** in September and last month relaunched a much improved version of the service. I asked Mark a few questions about fig**share** below. I also uploaded two datasetst to fig**share** and made them publicly available:

- CrowdoMeter Tweets – all 467 tweets used in the CrowdoMeter project
- CrowdoMeter Classifications – all 953 classifications from the CrowdoMeter project.

1. What is figshare?

figshare is a repository where users can make all of their research outputs available in a citable, sharable and discoverable manner. figshare allows users to upload any file format so that figures, datasets and media can be disseminated in a way that the current scholarly publishing model does not allow.

2. What is the right content for figshare? And what should rather go somewhere else?

Every experiment that is completed without error in the methods is valuable. Researchers investigate things because the question is interesting and supposedly unanswered. This means that other researchers will at some point ask that same question. Just because the hypothesis didn't turn out to be true, doesn't mean that this data should be thrown away.

With fig**share**, you can share your negative results or results that you were not planning to publish. You can make raw data available or supplementary material that journals cannot handle linked to from the published article. You can even make your papers and posters available and citable. People have uploaded whole chapters of their PhD thesis to share with the world and make sure that all of their hard work is not wasted.

Currently we are not focusing on handling massive datasets, whilst this is an aim for the future. These edgecases seem to be better handled by journals set

up specifically for publishing these huge files, such as the excellent GigaScience.

3. You relaunched figshare in January. What has changed to the previous version?

The old site was a proof of concept based on Mediawiki software. The new site has been completely built from scratch so that it is rapidly extendable in terms of features and scale. This means that we can adapt quickly and easily to the needs of researchers. As well as being much more intuitive, the biggest new feature for fig**share** is the private space.

Whilst we would like everyone to make all of their research objects available, we appreciate that some researchers would like to keep research private for many reasons. Because of this we set about giving users their own private repository to store their research objects. These objects can be uploaded in seconds and all objects are initially held in the private space, from where they can be made publicly available when the user decides. All research is easily tagged and categorisable, so that researchers can filter through their many files to find the one they were looking for in no time at all.

4. How important is the user interface for figshare? What particular features do you like the most?

The user interface is essential, researchers are busy enough as it it. They haven't got the time to attend training course on how to use a repository. If this was the case with facebook, no one would use facebook. For this reason fig**share** is stupidly simple and allows users to get their research onto the site in seconds, even the PIs. At this point they can choose to make it publicly available and immediately citable, sharable and discoverable, or keep it private – securely hosted, taggable and accessible when they need it, from anywhere in the world.

This is something we are constantly aiming to improve and we not only welcome feedback, we are actively seeking the thoughts of researchers on how we can make this a seemless part of their research process.

5. Do you have plans for a desktop version of figshare, e.g. to watch folders for new figshare content?

We do! By creating a desktop uploader, the process becomes even more intuitive for researchers, allowing them to make backups of their research in the cloud with no effort expenditure. Research data management is something I personally was terrible at. My research was organised into folders based on the month and the year I did that work. I lost days trying to find files that 'I know I worked on sometime last summer'. Hopefully a desktop uploader will add to the current simple management system so that researchers like me have no excuse when it comes to losing (often expensive) results files.

6. How is figshare different from data repositories?

fig**share** is not limited as many repositories are. It caters for all research domains, no matter where your research is carried out. Institutional repositories are often limited to members of said institution. Also, the majority of institutional repositories are built specifically for papers. The Dryad repository has been doing some great work making the datasets behind published articles available under CC0. fig**share** is not limited by the normal constraints of publishing, data generated in the lab can be shared and made available to the world as a citable object the same day.

fig**share** also gives the researchers the credit for their research. By adding metrics to the public uploads, and putting the cumulative metrics of a researcher's uploads on their profile, users can see the true impact and reach of the hard work they put in.

7. You currently use handles for figshare content. What are your thoughts on persistent identifiers? Are the plans to use DOIs?

Persistent identifiers are essential for the long term availability of research outputs. One of the reasons I set up figshare was because I wanted to cite a video in my thesis. Research data on YouTube is not easily citable, by adding persitent identifiers and an organised citation structure, videos as well as any other file format can be easily cited. All citations can be exported to Mendeley, Endnote and RefManwith one click. We use handles at the moment, but have noticed that researchers tend to be more familiar with DOI's and so will be making the move over to them shortly. We're currently working with DataCite through the British Library to get this set up.

8. How does figshare guarantee long-term preservation of uploaded data?

The long term persistence of this research data is essential. The research is backed up locally as soon as it is uploaded. We are currently in talks with Portico to back up all data and further guarantee this long term persistence.

9. What are your responsibilities at figshare? What did you do before Figshare?

I'm basically responsible for making the platform as useful for researchers as possible. I also love going to talk at Universities and conferences to researchers directly, to hear their thoughts/opinions. As a former life science researcher (I finished my PhD in stem cell biology at Imperial College London in September), I understand that what makes sense in a rational world does not make sense in the scientific world. You just have to look at the established model of scientific

research dissemination to understand that. fig**share** is useful whether you want to make all of your research outputs available or none. Science would just be a lot more efficient and dynamic if they did.