

# Open Research Pilot case studies: sharing image data

January 2017 saw the launch of the [Open Research Pilot Project](#). This two-year initiative comprised [four volunteer University of Cambridge research groups](#), University Research Support, and [Wellcome Trust's Open Research Team](#). The aims of the project were to perceive what is needed for researchers to make openly available, and be rewarded for, all outputs of the research process (e.g. along with traditional publications, other outputs include negative results, protocols, source code).

In the first of a series of blogs, Dr Ben Steventon talks to the research support team for the pilot about his group's involvement with the project. His particular Open interests through this project have been the how, where and what related to sharing very large file size image data.

## START OF PROJECT

Dr Ben Steventon applied to join the project because, while considerable advances have been made in the open sharing of sequencing data over the last ten years, the same cannot be said for the sharing of image data. He thinks that this is due to a few practical reasons that make it very time-consuming for researchers to get their imaging data ready for uploading to repositories.

Firstly, there is an increasing number of different imaging modalities available to researchers. While this is undoubtedly a very good thing for research, it does mean that it is very difficult to think of a standardized way to perform data annotation and descriptions of each image dataset. In itself, the fact that different labs will have different microscopy images of the same samples, or different samples with the same microscope is a good motivation for sharing image data in the first place. So much time is spent trying out different imaging set-ups at the start of a particular experiment and being able to access the trial-and-error periods of other research groups would be a major advantage to research. The process of going back over all the imaging data relating to a particular project or publication and annotating it in a way that is ready to share is a very time-intensive task. Perhaps the way forward would be to think of a way to integrate data collection with appropriate annotation as it is generated, so that the data would be ready to deposit at a click of a button. But how does one predict the specific repository at the start of a project that may last several years? In many cases there are no guarantees that data format and annotation templates will remain standard throughout this time.

A second barrier to data sharing is the overall size of imaging datasets, particularly now that image acquisition speed is increasing dramatically with new technologies such as light-sheet imaging. A single raw image data-set could exceed several terabytes. While some repositories are willing to take data of such size, it is not at all clear that researchers would be interested in starting with this level of the data. Many would be happy to start with already-processed imaging datasets, or even directly with the feature extracted and

analyzed data that come from them. How do we go about sharing data at multiple levels of analysis, when different repositories would be interested in only one or two of them? There has to be some centralized place for interested researchers to go to, and from there follow links to the various other places that host the data. Should such a website be hosted at the level of an individual lab, imaging facility, or should it be community based?

Ben saw the project as a chance to learn more from people within Cambridge University Libraries about the local resources available, and how they might be able to support the specific challenges outlined above. He was also interested to hear more from Wellcome Trust about new ways to share research data in a more general sense, in the hope that this would sharpen up his understanding of the challenges facing the open sharing of imaging data and how these might be overcome.

## **PROJECT IN PROGRESS**

Through the interactions with the library, Ben was introduced to a number of different repositories in the local area that might be able to host the data being generated in the lab. These interactions have focused his attention on thinking about a) how to build an integrated website that will allow researchers to link out to these various repositories and b) how to use tools such as electronic lab books to keep track of experimental and imaging information so that the required data annotated can be streamlined at the point of data deposition. Through interactions with Wellcome Trust, he has learnt of their researcher enhancement scheme for open research funding, and crucially about the specific things that they are looking for in such proposals. He feels in a much better position to take on the challenges outlined above and thinks that his team can make progress from this point on.

Recently, Ben has published [a paper in \*Development\*](#), that in part rests on results gained from a modified light-sheet microscope in the Cambridge Advanced Imaging Centre. Keen on sharing this data to a broader audience, he made use of both the [Image Data Repository to share the raw data](#) and a [Wellcome Open Research Data Note](#) to describing the data itself directly, taking it out of the context of the research article in *Development*. Making use of the Wellcome Open Research platform to publish data notes alongside resources such as IDR promises to be an effective manner of gaining increased feedback and audience for the data, as a series of additional reviewers are invited to peer-review the manuscript openly online. One difficulty of this approach was in coordinating the acceptance and publication of the data note and data with the research article. However, *Development* was very helpful in this respect as it was willing to add in the data sharing descriptions and citations as a correction post-publication.

The biggest surprise to Ben has been the degree of interest from Cambridge University Libraries and funders such as Wellcome Trust in trying to support Open Research in a broad sense. He observes that it seems to be a priority area for both groups, with researchers actually lagging behind a little in thinking about the problems associated with making this happen for the specific aspects relating to their research: clearly a lot more needs to be done to get these groups of people interacting more.

Over the course of the project, Ben has become much more aware of the specific problems involved relating to the open sharing of his research. He realises that the issues are going to be very field specific, but at the same time there will always be shared aspects between different fields. One positive aspect of the pilot was having researchers from these different fields together in the same room.

## **LOOKING AHEAD**

The principle thing Ben needs to get going is having a person in the lab with the appropriate skills and the time to focus on developing the framework required to have a sustainable system in place to share the team's research outputs. Fortunately, this has now been provided by enhancement funding from Wellcome Trust. The specific issues in terms of locating repositories and obtaining the appropriate support have been provided from his interactions with Cambridge University Libraries staff. He very much hopes that this relationship can continue.

Furthermore, he is very interested in working with the Cambridge Advanced Imaging Center in the hope that the framework that is developed in his lab can be expanded to other users of the facility. Having Open Research practices at the level of initial data acquisition and processing could be a very interesting way to move forward.

*As told to the Open Research Pilot Research Support Team*

*Published 5<sup>th</sup> February 2019 via <https://openadventures-blog.lib.cam.ac.uk/?p=197>*