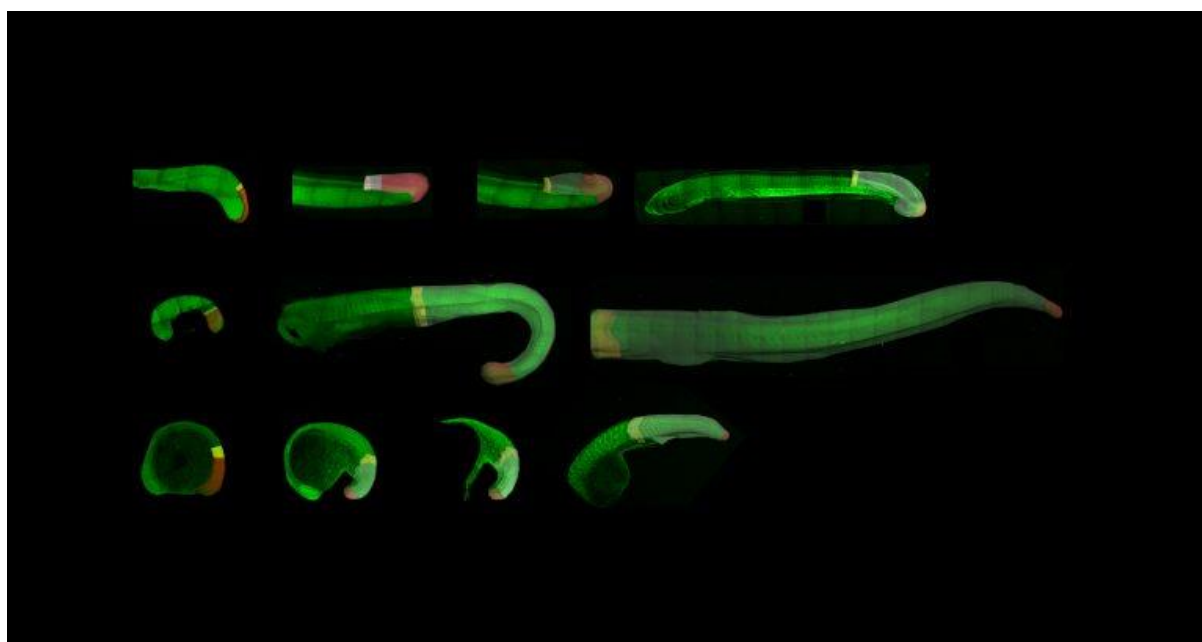


Open at scale: sharing images in the Open Research Pilot

Dr Ben Steventon is one of the participants in the [Open Research Pilot](#). He is working with the Office of Scholarly Communication to make his research process more open and here reports on some of the major challenges he perceives at the beginning of the project.

The [Steventon Group](#) is a new group established last year which looks at embryonic development, in particular focusing on the zebrafish. To investigate problems in this area the group uses time-lapse imaging and tracks cells in 3D visualisations which presents many challenges when it comes to data sharing, which they hope to address through the [Wellcome Trust Open Research Project](#). Whilst the difficulties that this group are facing are specific to a particular type of research, they highlight some common challenges across open research: sharing large files, dealing with proprietary software and joining up the different outputs of a group.

Sharing imaging data



The data created by time-lapse imaging and cell tracking is frequently on a scale that presents a technical, as well as financial, challenge. The raw data consists of several terabytes of film which is then compressed for analysis into 500GB files. These compressed files are of a high enough quality that they can be used for analysis but they are still not small enough that they can be easily shared. In addition the group also generates spreadsheets of tracking data, which can be easily shared but are meaningless without the original imaging files and specific software to allow the two pieces of data to be connected. One solution which we are considering is the [Image Data Resource](#), which is working to make imaging datasets in the life sciences, which have not previously been shareable due to their size, available to the scientific community to re-use.

Making it usable

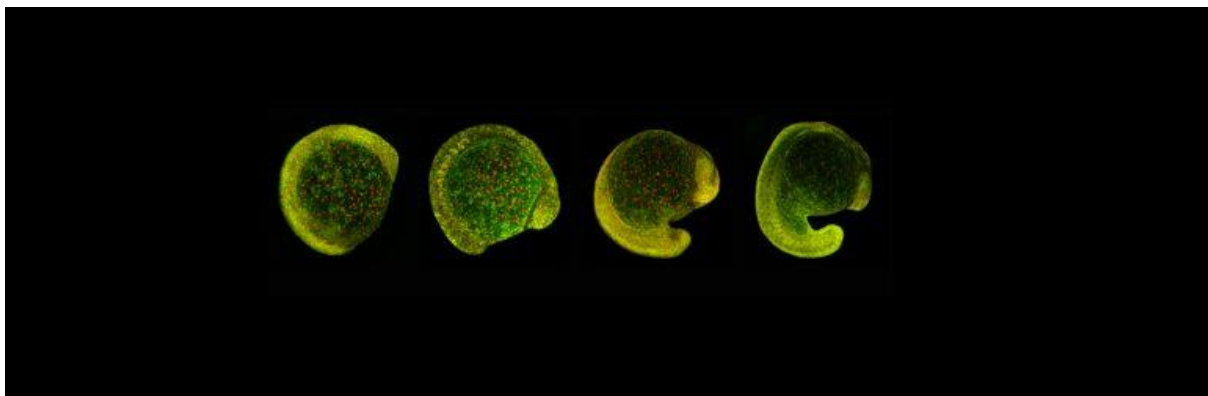


The software used in this type of research is a major barrier to making the group's work reproducible. The [Imaris](#) software the group uses costs thousands of pounds so anything shared in their proprietary formats are only accessible to an extremely small group of researchers at wealthier institutions, which is in direct opposition to the principles of Open Research. It is possible to use [Fiji](#), an open source alternative, to recreate tracking with the imaging files and tracking spreadsheets; however, the data annotation originally performed in Imaris will be lost when the images are not saved in the proprietary formats.

An additional problem in such analyses is the sharing of protocols that detail the methodologies applied, from the preparation of the samples all the way through data generation and analysis. This is a common problem with standard peer-review journals that are often limited in the space available for the description of methods. The group are exploring new ways to communicate their research protocols and have created an article for the [Journal of Visualised Experiments](#), but these are time consuming to create and so are not always possible. Open peer-review platforms potentially offer a solution to sharing detailed protocols in a more rapid manner, as do specialist platforms such as [Wellcome Open Research](#) and [Protocols.io](#).

Increasing efficiency by increasing openness

Whilst the file size and proprietary software in this type of research presents some barriers to sharing, there are also opportunities through sharing to improve practice across the community. Currently there are several different software packages being used for visualisation and tracking. Therefore, sharing more imaging data would allow groups to try out different types of images on different tools and make better purchasing decisions with their grant money. Furthermore, there is a great frustration in this area that lots of people are working on different algorithms for different datasets, so greater sharing of these algorithms could reduce the amount of time wasted creating algorithms when it might be possible to adapt a pre-existing one.



Shifting models of scholarly communication

As we move towards a model of greater openness, research groups are facing a new difficulty in working out how best to present their myriad outputs. The Steventon group intends to publish data (in some form), protocols and a preprint at the same time as submitting their papers to a traditional journal. This will make their work more reproducible, and it also allows researchers who are interested in different aspects of their work to access the bits that interest them. These outputs will link to one another, through citations, but this relies on close reading of the different outputs and checking references. The Steventon group would like to make the links between the different aspects of their work more obvious and browsable, so the context is clear to anyone interested in the lab's work. As the research of the group is so visual it would be appropriate to represent the different aspects of their work in a more appealing form than a list of links.

The Steventon lab is attempting to link and contextualise their work through their [website](#), and it is possible to cross-reference resources in many repositories (including Cambridge's [Apollo](#)), but they would like there to be a more sustainable solution. They work in areas with crossovers to other disciplines – some people may be interested in their methodologies, others the particular species they work on, and others still the particular developmental processes they are researching. There are opportunities here for openness to increase the discoverability of interdisciplinary research and we will be exploring this, as well as the issues around sharing images and proprietary software, as part of the Open Research Pilot.

Published 8 May 2017 via <https://unlockingresearch-blog.lib.cam.ac.uk/?p=1444>

Written by Rosie Higman and Dr Ben Steventon