

Images to reveal all?

It's a common frustration that most readers have probably experienced at one time or another: you see an intriguing image in a paper, but want to know just that bit more than is described. Just how bright is bright: what were the microscope settings when the image was taken? How many sections were used to demonstrate the apparent colocalization of two proteins? Where is the particle moving between two time-points?

Our increasing reliance on live imaging analysis for cell biological research and the digital processing that most imaging data undergo before publication brings this issue to the forefront of our minds (also see Rossner, M. & Yamada, K. *J. Cell Biol.* 166, 11; 2004). Is enough information provided for all readers to be able to independently assess the published data? If not, then is it reasonable for reviewers and/or readers to ask to see the unprocessed datasets? Of course, everyone relies heavily on each other's professional integrity and it is generally assumed that the data shown is that which is described. But even in an ideal world in which this were always true, we may still be left with the sub-optimal situation in which other laboratories find it difficult to reproduce experiments because key details of image acquisition are not provided.

A practical limitation is the format constraints of the printed page. The advent of online-only data that accompanies papers has solved this space problem to some extent. One possibility, then, is that researchers provide the unprocessed imaging data to accompany a manuscript online. It is equally important that details of how the images were recorded be provided. To this end, we do not have a formal word limit for the methods section of this journal other than a restriction to essential information that is sufficient to reproduce the data represented.

Perhaps a better solution would be for all of the details about an image and how it was acquired to remain part of the image file itself. A significant step towards this can be found in new databases that are developing common formats for describing image data. The Open Microscopy Environment (OME; www.openmicroscopy.org), for example, has been set up in response to the realization that: "no universally accepted formats exist for storing and sharing multi-dimensional image data (having spatial, temporal and spectral range) and for preserving the all-critical metadata that records what the experiment is about. Even more importantly, relatively few analytical routines are available for extracting quantitative data from images and linking them to other biological information". This initiative, founded by Jason Swedlow, Peter Sorger and Ilya Goldberg, is an open-source software project that aims to develop a database-driven system for the quantitative analysis of biological images.

But an important aspect of this project is that it provides a logical system by which to describe an image dataset. All datasets can be stored in an XML file format that should be easily transferable

between researchers, and will include the metadata, that is, all the details of the image's history. In this way, it provides a permanent fingerprint of the experiment. This file format can include three-, four- and five-dimensional (5D) images, as well as the all-important information about how the microscope was set up, from the light-sources being used to the filters and detector settings. It will also be able to store within one file a library of single-plane images, so that researchers can view the data for themselves section by section.

So, although the main aim of the OME is to allow an open-source environment to develop software for analysing data, a vital by-product is that it has also defined "a set of information common to all biological microscopy, a format for storing this information, and a mechanism for sharing it with other software". In theory, this type of file format could be adopted by researchers for storing their image data, and could be used to transfer files between labs.

Assuming we agree it is reasonable that this data be made accessible, the next question is where all this data could be stored. File size is obviously a concern once you get into 4D datasets, but one advantage of XML files is that they are easily compressible. So, should researchers be expected to make such files available on request after publication, or is this the responsibility of the journals in which the work is published? On the basis of an initial survey of the community, it seems there are mixed feelings about this.

Another possibility is that images be placed in an independent image database, akin to the protein data bank (www.rcsb.org/pdb). One initiative that aims to provide such a repository is the BioImage database (www.bioimage.org). This project, headed by David Shotton at Oxford University, is being developed with the idea that it is "essential to acquire not just high-quality source images, but also the various images derived from them, and detailed metadata documenting the process of their creation" (*Nature* 422, 374; 2003). BioImage plans to provide a public resource of both published and unpublished images, together with the detailed metadata documenting what they are and how they were created, which can then be researched by the community. It can also accept and publish images created using OME. Of course, it would be imperative that any database of this type be freely available, as well as permanent and citable.

At *Nature Cell Biology*, we support the development of databases that fit the above criteria and that are endorsed by the community, as well as the adoption of standard file formats that allow easy access to the unprocessed image data and the vital information about how this was obtained. Of course, such moves depend on the standards that researchers in the field think it is reasonable to hold each other to. At a minimum, we feel that reviewers should be able to request unprocessed imaging data wherever necessary to judge the data presented, and that authors should provide information about images in sufficient detail to allow reproduction. We are continuing to actively explore this issue and require your feedback as to how we can best provide you access to imaging data that goes beyond a teasing snapshot.