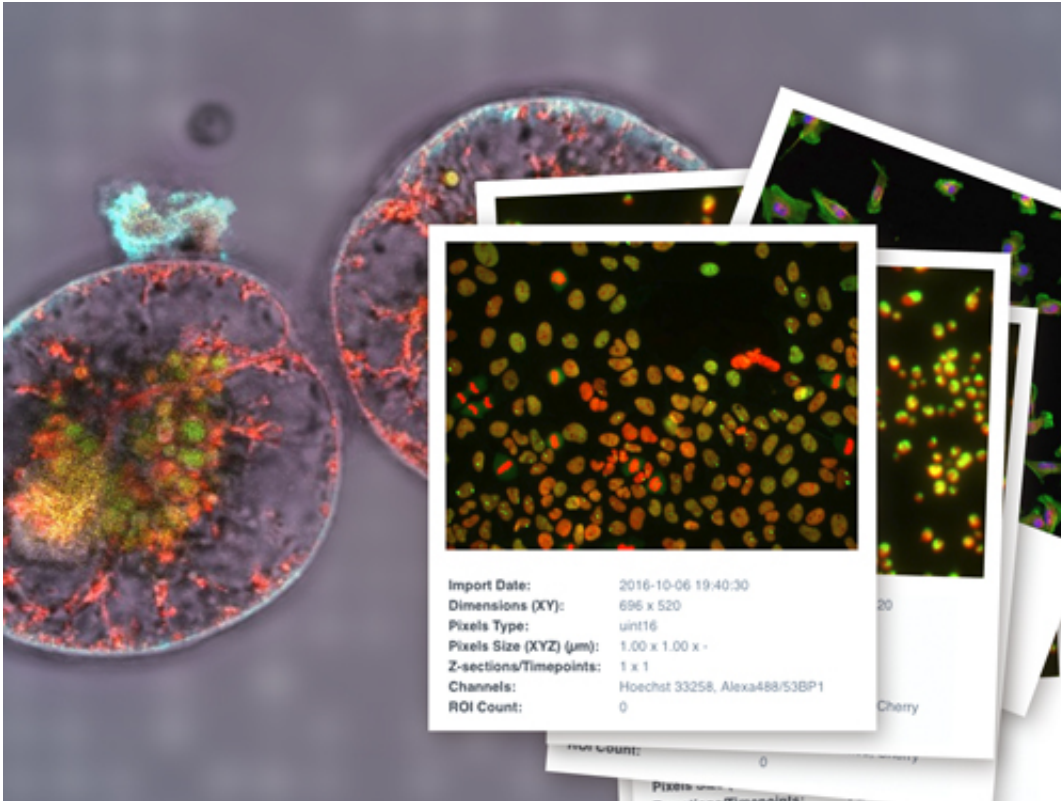


Open imaging data for biology

June 21 2017



Credit: European Bioinformatics Institute EMBL-EBI

A picture may be worth a thousand words, but only if you understand what you are looking at. The life sciences rely increasingly on 2-D, 3-D and 4-D image data, but its staggering heterogeneity and size make it extremely difficult to collate into a central resource, link to other data types and share with the research community.

To address this challenge, scientists at the University of Dundee, the European Bioinformatics Institute (EMBL-EBI), the University of Bristol and the University of Cambridge have launched a prototype repository for imaging data: the Image Data Resource (IDR). This free resource, described in Nature Methods, is the first general biological image repository that stores and integrates data from multiple modalities and laboratories.

The IDR also reveals the potential impact of sharing and reusing imaging data for the life sciences.

Pooling resources

"Imaging will only be truly transformative for science if we make the data publicly available," explains Alvis Brazma, a lead author and Senior Scientist at EMBL-EBI. "Scientists should be able to query existing data to identify commonalities and patterns. But to make this possible we need a robust platform where researchers can upload their imaging data and easily access data from other experiments. The Image Data Resource is the first step towards creating a public image data repository for the [life sciences](#)."

There are many resources worldwide in which people publish imaging data, but none of these repositories is both generic and linked to other relevant bio-molecular data. This means that for all the effort that goes into them, it is difficult to reuse these datasets in new studies.

There are many reasons why sharing imaging data has been so difficult until now – most notably the heterogeneity and complexity of the image data, but also a critical mass of storage, compute and curation expertise.

"Imaging data is large, yes, but the real challenge is that it is heterogeneous and multidimensional," says Jason Swedlow, senior author

of the study and Professor of Quantitative Cell Biology at the University of Dundee. "Curating, storing and analysing imaging data require significant effort and computing power. The creation of the IDR prototype was only possible thanks to a strong collaboration between several scientific organisations."

Nice picture – but what does it mean?

IDR contains a broad range of imaging data, including high-content screening, super-resolution microscopy, time-lapse and digital pathology imaging. But it's not just the diversity of data types that makes the resource unique; it is the additional information available that creates the added value.

"IDR doesn't just show you an image or a video of a cell. It also tells you what the image is about, where it was taken, by whom and what conclusions can be drawn from it," continues Brazma.

The new resource integrates imaging data with molecular and phenotype data. IDR includes information on experimental protocols: parameters, analyses and the effects scientists have observed in cells and features, for example. This makes it possible for users to analyse gene networks – potentially revealing previously unknown interactions – on a scale that would not be possible for individual studies. That requires a staggering amount of storage and compute power. The IDR collaboration was able to launch their project successfully thanks to the Embassy Cloud resource and support at EMBL-EBI.

The Image Data Repository

The prototype public image repository contains a broad range of data, including:

- High-content screening
- Super-resolution microscopy
- Time-lapse imaging
- Digital pathology imaging
- Experimental protocol metadata
- Observed effects in cells and features
- Cross references with molecular archives

Demonstrating success

The Swedlow group at Dundee and the Carazo Salas group at the University of Bristol used IDR to illustrate how shared imaging data can push the boundaries of research. Using data deposited in the IDR, they identified genes from different studies that, when mutated or removed, caused cells to elongate and stretch out. They put together information from several different studies and built a gene network, which gives a clear view on how these genes affect cell shape – an important property to consider in metastatic cancer.

"Expanding the public archives to include imaging is of huge interest to the biotech industry and drug development companies. It offers potential to identify new therapies and targets, and broadens the scope of research by allowing scientists around the world to access each other's imaging datasets," adds Swedlow.

"Bioimaging technologies are currently revolutionising life science. Sharing the rapidly increasing amount of image data is the key to enabling ground-breaking future research," says Jan Ellenberg, Head of EMBL's Cell Biology and Biophysics Unit and Coordinator of Euro-BioImaging, the pan-European infrastructure for imaging technologies. "For this reason image data archiving and sharing is a high priority for EMBL, and for Euro-BioImaging's future general data services, which can build on the IDR pilot example."

Next steps

So far, the collaborators have proven that IDR is both possible and useful. The next step is to secure the support and investment needed to transform the prototype into a production-ready imaging infrastructure.

IDR's software and technology is open source, so it can be accessed and built into other [image data](#) publication systems. This promotes and extends publication and re-analysis of scientific data.

More information: Eleanor Williams et al. Image Data Resource: a bioimage data integration and publication platform, *Nature Methods* (2017). [DOI: 10.1038/nmeth.4326](https://doi.org/10.1038/nmeth.4326)

Provided by European Bioinformatics Institute EMBL-EBI

Citation: Open imaging data for biology (2017, June 21) retrieved 25 April 2024 from <https://phys.org/news/2017-06-imaging-biology.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.