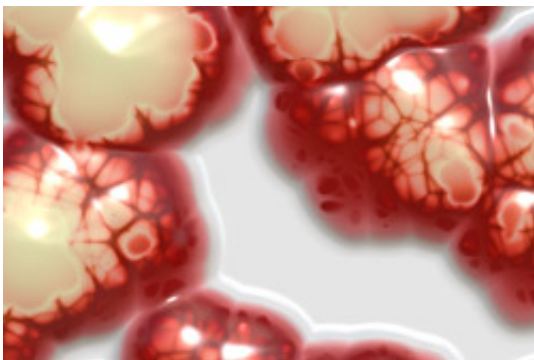


'Big Data' resource raises possibility of research revolution

June 13 2017



Credit: University of Bristol

A group of UK scientists involving researchers from the University of Bristol have demonstrated how aggregating image data from laboratories all around the world has the potential to revolutionise scientific research.

The Image Data Resource (IDR) is a collaboration between [scientists](#) in the Open Microscopy Environment (OME), based at Dundee, and groups at Universities of Cambridge and Bristol, and the European Bioinformatics Institute (EMBL-EBI). The collaboration brings together biologists, imaging specialists, big data scientists and computer scientists.

A team headed by Professor Jason Swedlow in Dundee's School of Life Sciences has built the Image Data Resource (IDR), a public database that

collects and integrates imaging data related to experiments published in leading scientific journals. This means that 'Big Data' from imaging experiments conducted by scientists all over the world that were previously too large and difficult to share are now publicly available.

Access to primary research data is vital for the advancement of science but comparing and analysing image datasets produced by individual researchers is notoriously difficult for scientists. The [images](#) are large, unwieldy, complex and heterogeneous. They are rarely publicly available and, even if they are, different means of collating and storing image data mean they cannot be easily reproduced, compared or re-analysed.

IDR automates these processes and pulls individual pieces of related research together to create a vast bank of knowledge that can save researchers time, effort and money while serendipitously highlighting previously unexplored areas with the potential to solve scientific mysteries. This free resource is the first general biological image repository that stores and integrates data from multiple modalities and laboratories.

Professor Swedlow explained, "Researchers collaborate with each other and keep abreast with research work from the global scientific community at meetings and in published papers, but the image datasets that underpin these communications are almost never published. As a result there is a huge amount of information that cannot be shared, accessed, compared or understood.

"IDR makes these datasets available, and allows scientists worldwide to combine, mine and analyse these imaging data. The potential to speed up research and link datasets so that scientists can look for patterns and commonalities is enormous. Even before officially announcing IDR, we've had contacts from cell biologists, drug discovery scientists and deep learning developers asking if they can use IDR."

IDR collects and integrates imaging data acquired across many different imaging modalities. It links high-content screening, super-resolution microscopy, time-lapse and digital pathology imaging experiments to public genetic or chemical databases. IDR also includes information on experimental protocols, imaging parameters, analyses and the cellular and tissue changes scientists have observed.

Using IDR, Professor Swedlow and his colleagues in the Open Microscopy Environment (OME) Consortium found connections between different research projects that had eluded individual researchers. They identified genes from different studies that, when mutated or removed, caused cells to elongate and stretch out.

They assembled gene lists from the different studies and built a gene network that gives a more complete picture of how genes control shape, one of the properties that change in metastatic cancer. Elongation is just one of more than 150 effects on cells that IDR currently records, meaning further significant discoveries are anticipated.

This area is of huge interest to the biotech industry and drug discovery companies because of its potential to identify new therapies and targets and broaden the scope of research by allowing scientists to access each other's datasets.

"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."

Professor Rafael Carazo Salas, who led the IDR team at Cambridge and Bristol, said, "Reproducibility and re-use are key concerns in the scientific community. We have shown how they can enhance research by integrating and cross-validating different imaging studies and by enabling the generation of discoveries, added value and increased return on investment that could not be obtained from individual studies on their own.

"IDR is world-leading not only because of the discoveries it makes for the first time possible but also because it is an open source platform that others can use to publish their own image data. Thus IDR provides both a novel on-line resource and a software infrastructure that promotes and extends publication and re-analysis of scientific [image data](#)."

Provided by University of Bristol

Citation: 'Big Data' resource raises possibility of research revolution (2017, June 13) retrieved 3 May 2024 from <https://phys.org/news/2017-06-big-resource-possibility-revolution.html>

<p>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.</p>
--