IMAGING TECHNOLOGY

Beautiful BioImages for the Eyes of Many Beholders

A handful of image-sharing databases and software systems is becoming available, and these projects might change the way biologists look at their own and other researchers' data

For some biologists, seeing is believing. They apply fluorescent tags to track proteins, use multidimensional microscopes to watch embryonic development unfold, or spy on endangered species' mating habits using video. Yet, only a fraction of these images finds its way into publications, and then only in a static, two-dimensional format.

Many labs have treasure chests of images never seen by colleagues, and researchers are now trying to find ways to share the wealth. It's not strictly a philanthropic impulse: Sharing images would save time, allowing researchers to compare and build on one another's findings. It might even lead to research projects that would be impractical for just one lab. "In today's lab meeting, someone showed an image—[the person] had spent 8 hours at the microscope just collecting the data on this one image," explains Scott Fraser, director of the Biological Imaging Center at the Beckman Institute of the California Institute of Technology (Caltech) in Pasadena. "It would save others a lot of time if [those] data could be used more than once."

Several projects now coming online aim to make images accessible to all. To do this, researchers are developing new ways to store, link, search, and retrieve biological data. The situation today is like "a whole bunch of blind people all feeling the same elephant—and we don't really realize what the others are about," says G. Allan Johnson, director of the Center for In Vivo Microscopy at Duke University. "The elephant we are all connected to is biology."

The obstacles these projects face loom large. Aside from the technical difficulties of creating user-friendly databases and interconnected networks of images in the scientific literature, there are pesky legal and ethical questions. It's not clear who will own the copyright to images in a database. And how do you give credit to the people who generate an image if your work builds on theirs? "It is easy for someone ancient like me to be generous, because I do not have to worry about getting the next job, but the younger scientists are concerned," says Fraser.

The big picture

A project called BioImage, funded by the European Commission (EC), leads the effort to store and mine images in the scientific literature. It was thought up in the mid-1990s by a handful of scientists in Germany, Switzerland, the United Kingdom, and Spain; the project is now run mainly by cell biologist David Shotton of Oxford University. Together with Oxford's Steffen Lindek, a consulting biophysicist, and other colleagues, Shotton is developing what will be a Web-accessible database including everything from three-dimensional confocal microscopy images to time-lapse videos of cells to wildlife photos and videos. The goal of the project is to develop sophisticated image bioinformatics that will, in Shotton's words, "ensure that our image repositories become knowledge resources rather than data graveyards."

Who's there? Image-sharing projects will make candid wildlife shots available to conservationists and pattern-recognition researchers alike.

CREDIT: GREVY'S ZEBRA SURVEY 2000

BioImage, which is part of an EC research initiative called Online Research Information Environment for the Life Sciences, obtained a 3-year, 3 million Euro grant in January. BioImage builds on a prototype image database created by a related EC project. Now, Shotten and others are improving user interfaces and gutting the scaffolding of the database. By September, they expect, they will be able to incorporate not only microscopy images but also wildlife photographs and videos.

The first contributions to the new database will be images from the *Journal of Microscopy*. BioImage set up an agreement with the journal to bank the images associated with biological papers accepted for publication, including some images that won't appear in the journal. "We hope that a culture will develop for submission of images to the BioImage database, [one] that resembles the present culture of submission of sequence data and crystallographic information to the public databases," Shotton says.

BioImage won't warehouse all the images, though. In many cases, it will store descriptions and low-resolution visual previews and will direct searchers to sites where the high-resolution versions are parked. Most journals are expected to keep copyrighted images on their own servers rather than storing copies in the BioImage database.

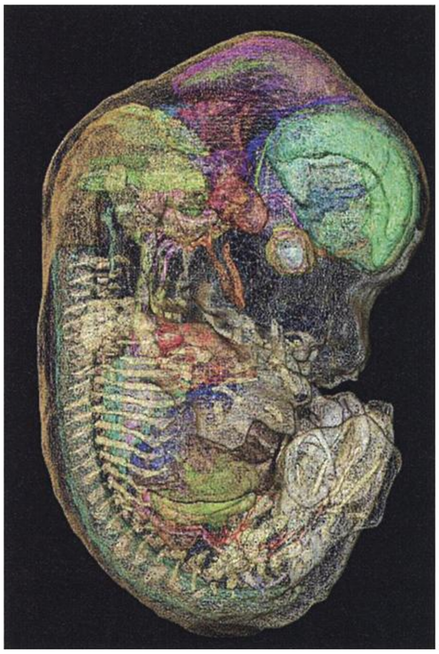
Because images will need searchable tags, they will be equipped with metadata such as a list of authors, a description of the experiment or observation, and details such as the microscope's aperture and exposure. BioImage will hold the copyright to the metadata, but the image copyright may be elsewhere—for example, with a publisher. The metadata and images stored in BioImage will be freely available for noncommercial research and educational uses. It's not yet clear what fees commercial researchers will have to pay. And all users will have to contact original copyright holders for images not kept in the database. BioImage researchers are also hoping to convince publishers to add tags to individual images within papers, to facilitate image searches.

Lindek, who developed the database software, foresees a time when researchers will do new virtual experiments using banked images. In accessing raw three-dimensional data from a confocal image, for instance, a researcher might download a stored image and reevaluate it from a new perspective or apply a different algorithm, perhaps obtaining publishable results. "In this sense, BioImage may allow a new kind of science to occur," says Shotton.

Professor Picasso's fluorescent period

BioImage has a broad mission, but other groups are setting up databases with more limited objectives. In the United States, the National Institutes of Health is sponsoring several focused projects, including one looking at genes active in a mouse's nervous system and another looking at slices of human brains. Another effort tracks cells as they move.

To explore how the nervous system develops, Nathaniel Heintz and Marybeth Hatten of Rockefeller University in New York City, along with Alexandra Joyner at New York University, are creating thousands of transgenic mice. They plan to insert fluorescent markers into genes. Then, they will capture images showing when and where the genes are active in neurons at embryonic and postnatal stages of growth. The tag they're using, green fluorescent protein, makes lovely images: "With this marker you can see the morphology of each neuron and tell the kind of neuron that is developing," explains Joyner.



Technicolor trajectories. Studies of genes active during development will illuminate images of embryonic mice.

This project, dubbed the Gene Expression Nervous System Atlas, will expand the number of people with access to these expensive, high-resolution pictures. Hosted by the National Center for Biotechnology Information, the database is expected to come online this year.

Another collection will hold high-resolution images, at a range of scales and collected using a variety of methods, of human brains. Work on the Biomedical Informatics Research Network (BIRN) began last fall. The initiative, which focuses on brain disorders, started with a $20 million grant to Duke University, Harvard, Caltech, the University of California (UC), San Diego, and UCLA, to create an information infrastructure that permits large-scale data exchange. "Neuroscientists studying such devastating diseases as Parkinson's disease and schizophrenia [will be able] to integrate information ranging in scale from the whole brain down to a single neuron," explains Duke's Johnson, a principal investigator of a related project called the Mouse Brain Imaging Research Network (MBIRN).

For both BIRN and MBIRN, which will focus on mouse models of human neurologic diseases, researchers will store images at their own institutions. But their troves will be linked by a so-called storage research broker, a program currently being developed at the UC San Diego Supercomputer Center. Much as in BioImage, a query will take a researcher to wherever the images are, in something of a scientific equivalent of an Internet search engine.

Right now, BIRN scientists are hammering out ways to manage these specialized databases and their connections. As Johnson explains, they're trying to develop the infrastructure with a constant eye toward biological research, basing their networks on, for example, a particular mouse model of a human disease that a lot of researchers are using. Says Johnson, "We [hope we] have the right collection of neuroscientists and geeks pounding on the same problem."

Meanwhile, another group is developing a small software project with a large scope. Open Microscopy Environment (OME) is the brainchild of three scientists: Ilya Goldberg and Peter Sorger of the Massachusetts Institute of Technology and Jason Swedlow of the Wellcome Trust Biocentre at the University of Dundee, U.K. This project will allow researchers to keep track of data that amass during observations of cells and their behavior. The team's open-source software, now in beta testing, is free to download.

Pointing to a blue and green glowing, diamond-shaped structure on his computer screen, which depicts a cell going through mitosis, Swedlow says, "Just about anyone can fix a cell, label what is of interest, and create these kinds of images. Even labeling a protein in live cells and getting those images is not really the problem anymore."

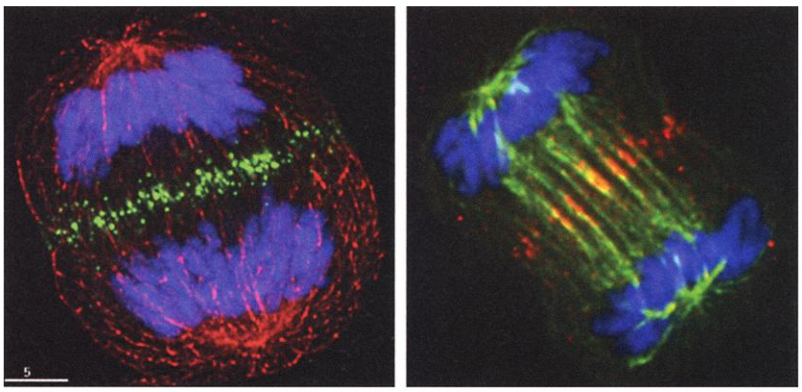
What is now of interest, explains Swedlow, is to go beyond the pretty pictures to probe the quantitative information in the image. For example, it is routine to record time-lapse movies of many cells at once, by quickly focusing and refocusing on different ones. A researcher might be looking at how different drugs or different mutations affect the cells. The challenge, Swedlow says, is "finding what has changed, what has moved, measure in what way fluorescence has shifted, and evaluate those differences."

Popular algorithms are now available for analyzing images, but the summary results—the metadata—tend to get buried by technological change. Each time a researcher uses a different software package, some of the old data are left behind. There are no accepted standards for storing the image data, a basic problem that, Swedlow says, "we want to solve." OME establishes a local database management system that stores both the data and metadata. "By preserving the relationships between them, it becomes easier to evaluate large numbers of images."

Like BioImage and BIRN, OME draws on the concept of federated databases, enabling information from several sources to be linked and exploited. "Creating these links is the only way to allow this collection to be queried for a search," explains Swedlow. This and other new strategies might soon be ways to raise biological images from obscurity back into the realm of the visible and findable.

–VIVIEN MARX

Vivien Marx is a freelance writer based in Boston, Massachusetts, and Frankfurt, Germany. Her latest book is *The Semen Book*.



Fleeting data. Sophisticated software is necessary for quantifying and keeping track of the changes visible as cells divide.

CREDITS: (TOP) SETH RUFFINS/CALTECH; (BOTTOM) PAUL ANDREWS AND JASON SWEDLOW/UNIVERSITY OF DUNDEE