

data on spatiotemporal dynamics of signaling reactions will require the development of new analysis methods, and the assembly of the data in conceptual frameworks will require mathematical modeling (7). It is to be expected that the ability to measure signaling in individual cells, which up to now has mostly taken place in isolated cells in tissue culture, will become increasingly important in the analysis of signaling reactions in tissues and embryos. Such measurement could revolutionize the study and understanding of signaling during development.

References and Notes

1. A. G. Gilman *et al.*, *Nature* **420**, 703 (2002).
2. W. Wallace, L. H. Schaefer, J. R. Swedlow, *Biotechniques* **31**, 1076 (2001).
3. D. Toomre, D. J. Manstein, *Trends Cell Biol.* **11**, 238 (2001).
4. Y. Harada *et al.*, *Methods Cell Biol.* **55**, 117 (1998).
5. S. R. Neves, R. Iyengar, *Bioessays* **24**, 1110 (2002).
6. F. Castiglione, M. Bernaschi, S. Succi, R. Heinrich, M. W. Kirschner, *Phys. Rev. E* **66**, 031905 (2002).
7. R. D. Phair, T. Misteli, *Nature Rev. Mol. Cell Biol.* **2**, 898 (2001).
8. L. Jaffe, *Cell Calcium* **32**, 217 (2002).
9. J. C. Gilkey, L. F. Jaffe, E. B. Ridgway, G. T. Reynolds, *J. Cell Biol.* **76**, 448 (1978).
10. N. Dumollard, J. Carroll, G. Dupont, C. Sardet, *J. Cell Sci.* **115**, 3557 (2002).
11. L. L. Runft, L. A. Jaffe, L. M. Mehlmann, *Dev. Biol.* **245**, 237 (2002).
12. M. J. Berridge, P. Lipp, M. D. Bootman, *Nature Rev. Mol. Cell Biol.* **1**, 11 (2000).
13. M. D. Bootman, M. J. Berridge, H. L. Roderick, *Curr. Biol.* **12**, R563 (2002).
14. S. Schuster, M. Marhl, T. Hofer, *Eur. J. Biochem.* **269**, 1333 (2002).
15. S. Shimozono *et al.*, *Sci. STKE* **2002**, pl4 (2002), http://stke.sciencemag.org/cgi/content/full/OC_sigtrans;2002/125/pl4.
16. P. W. Marks, F. R. Maxfield, *J. Cell Biol.* **110**, 43 (1990).
17. A. L. Kindzelskii, H. R. Petty, *J. Immunol.* **170**, 64 (2003).
18. ———, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9207 (2002).
19. L. F. Olsen, U. Kummer, A. L. Kindzelskii, H. R. Petty, *Biophys. J.* **84**, 69 (2003).
20. T. C. Rich, J. W. Karpen, *Ann. Biomed. Eng.* **30**, 1088 (2002).
21. A. Honda *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2437 (2001).
22. M. Zaccolo *et al.*, *Nature Cell Biol.* **2**, 25 (2000).
23. M. Zaccolo, L. Filippin, P. Magalhaes, T. Pozzan, *Novartis Found. Symp.* **239**, 85 (2001).
24. R. A. Firtel, C. Y. Chung, *Bioessays* **22**, 603 (2000).
25. M. Iijima, Y. E. Huang, P. Devreotes, *Dev. Cell* **3**, 469 (2002).
26. J. M. Haugh, F. Codazzi, M. Teruel, T. Meyer, *J. Cell Biol.* **151**, 1269 (2000).
27. A. Hall, C. D. Nobes, *Philos. Trans. R. Soc. London B* **355**, 965 (2000).
28. S. Eden, R. Rohatgi, A. V. Podtelejnikov, M. Mann, M. W. Kirschner, *Nature* **418**, 790 (2002).
29. O. D. Weiner *et al.*, *Nature Cell Biol.* **4**, 509 (2002).
30. P. A. Iglesias, A. Levchenko, *Sci. STKE* **2002**, re12 (2002), <http://stke.sciencemag.org/cgi/content/full/sigtrans;2002/148/re12>.
31. H. Meinhardt, *Bioessays* **22**, 1048 (2000).
32. D. Dormann, G. Weijer, C. A. Parent, P. N. Devreotes, C. J. Weijer, *Curr. Biol.* **12**, 1178 (2002).
33. E. M. Gardiner *et al.*, *Curr. Biol.* **12**, 2029 (2002).
34. N. Mochizuki *et al.*, *Nature* **411**, 1065 (2001).
35. T. Balla, P. Varnai, *Sci. STKE* **2002**, pl3 (2002), <http://stke.sciencemag.org/cgi/content/full/sigtrans;2002/125/pl3>.
36. S. Dowler, G. Kular, D. R. Alessi, *Sci. STKE* **2002**, pl6 (2002), <http://stke.sciencemag.org/cgi/content/full/sigtrans;2002/129/pl6>.
37. D. Dormann, T. Libotte, C. J. Weijer, T. Bretschneider, *Cell Motil. Cytoskeleton* **52**, 221 (2002).
38. C. D. Ellison *et al.*, *Curr. Biol.* **11**, 1631 (2001).
39. R. Keller, *Science* **298**, 1950 (2002).
40. P. N. Adler, *Dev. Cell* **2**, 525 (2002).
41. J. B. Wallingford *et al.*, *Mol. Biol. Cell* **11**, 769 (2000).
42. D. E. Levy, J. E. Darnell Jr., *Nature Rev. Mol. Cell Biol.* **3**, 651 (2002).
43. S. X. Hou, Z. Zheng, X. Chen, N. Perrimon, *Dev. Cell* **3**, 765 (2002).
44. T. Kawata *et al.*, *Cell* **89**, 909 (1997).
45. T. Araki *et al.*, *EMBO J.* **17**, 4018 (1998).
46. D. Dormann, T. Abe, C. J. Weijer, J. Williams, *Development* **128**, 1081 (2001).
47. M. Bienz, *Nature Rev. Mol. Cell Biol.* **3**, 328 (2002).
48. K. M. Pang, E. Lee, D. A. Knecht, *Curr. Biol.* **8**, 405 (1998).
49. J. V. Small, I. Kaverina, *Curr. Opin. Cell Biol.* **15**, 40 (2003).
50. C. M. Waterman-Storer, G. Danuser, *Curr. Biol.* **12**, R633 (2002).
51. N. Watanabe, T. J. Mitchison, *Science* **295**, 1083 (2002).
52. M. G. Vicker, *FEBS Lett.* **510**, 5 (2002).
53. W. M. Shih, Z. Gryczynski, J. R. Lakowicz, J. A. Spudich, *Cell* **102**, 683 (2000).
54. J. Zhang, R. E. Campbell, A. Y. Ting, R. Y. Tsien, *Nature Rev. Mol. Cell Biol.* **3**, 906 (2002).
55. F. Gaits, K. Hahn, *Sci. STKE* **2003**, pe3 (2003), <http://stke.sciencemag.org/cgi/content/full/sigtrans;2003/165/pe3>.
56. K. M. Pang, E. Lee, D. A. Knecht, *Mol. Biol. Cell* **8**, 964 (1997).
57. The data in Figs. 1, 2, and 3 were from experiments performed by D. Dormann (Univ. of Dundee). The statA GFP translocation experiments were performed as a collaboration with T. Abe and J. Williams (Univ. of Dundee). The GFP-tagged actin-binding domain of *Dictyostelium* ABP120 was a gift from D. Knecht (Univ. of Connecticut). I thank the Wellcome Trust and BBSRC for financial support.

Supporting Online Material
www.sciencemag.org/cgi/content/full/300/5616/96/DC1
 Movies S1 to S4

VIEWPOINT

Informatics and Quantitative Analysis in Biological Imaging

Jason R. Swedlow,^{1*} Ilya Goldberg,² Erik Brauner,³ Peter K. Sorger^{3,4}

Biological imaging is now a quantitative technique for probing cellular structure and dynamics and is increasingly used for cell-based screens. However, the bioinformatics tools required for hypothesis-driven analysis of digital images are still immature. We are developing the Open Microscopy Environment (OME) as an informatics solution for the storage and analysis of optical microscope image data. OME aims to automate image analysis, modeling, and mining of large sets of images and specifies a flexible data model, a relational database, and an XML-encoded file standard that is usable by potentially any software tool. With this design, OME provides a first step toward biological image informatics.

Recent excitement in optical microscopy centers on the extraction of quantitative numerical information from digital images to generate and test specific scientific hypotheses. For example, combining computer vision and speckle microscopy makes it possible to test specific mechanistic models of actin flow during cell movement (1). The potential for

automation in digital imaging is also driving interest in the use of microscopy for large-scale "screening by imaging," in which cells or organisms are treated with libraries of small molecules, banks of small inhibitory RNAs, etc. to identify chemicals or genes that affect a particular biological process by virtue of a change in cellular behavior or appearance

(2, 3) (Fig. 1). However, the routine application of automated image analysis and large-scale screening is held back by substantial limitations in the software used to store, process, and analyze the large volumes of information generated by digital imaging. It is possible to interpret

¹Division of Gene Regulation and Expression, Wellcome Trust Biocentre, University of Dundee, Dow Street, Dundee DD1 5EH, Scotland. ²Laboratory of Genetics, National Institute on Aging, National Institutes of Health, 333 Cassell Drive, Suite 4000, Baltimore, MD 21224, USA. ³Institute of Chemistry and Cell Biology, Harvard Medical School, 250 Longwood Avenue, Boston, MA 02115, USA. ⁴Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed. E-mail: jswedlow@dundee.ac.uk

images only if we know the context in which they were acquired. Current software for microscopy automates image acquisition and provides hardware and software solutions for three-dimensional (3D) imaging (using deconvolution and confocal and other methods) but does not keep track of image and analytical data in a rigorous way. It is usually possible to specify file name, date, and experimenter, but few packages systematically record the identities of the genes being studied, the labels used, etc. (4). Interoperability between different software systems involves the exchange of TIFF files, which preserve none of the contextual information. In this Viewpoint, we describe the conceptual challenges faced by image informatics as applied to biological microscopy and describe some of the solutions incorporated in an open-source image informatics system currently under development in our laboratories: the Open Microscopy Environment (OME) (5).

The primary goal of OME is to enable the automatic analysis, modeling, and mining of large image sets with reference to specific biological hypotheses. OME aims to manage images from all optical microscopes, including confocal, wide-field, and multiphoton systems, but other image types (such as computed tomography scans) are not necessarily supported. OME also aims to store, without loss or degradation, primary image data and the metadata that specify the context and meaning of an image. Some metadata are devoted to describing the optics of the microscope, some to the experimental setup and sample, and some to information derived by analysis. Finally, OME aims to provide a flexible mechanism for incorporating new and existing image analysis routines and storing the output of those routines in a self-consistent and accessible manner.

The OME Data Model and Database

The OME data model is a formal description of the structure, meaning, and behavior of data stored and manipulated by the system, and is instantiated via both a database and a file format (Fig. 2). The OME data model has three parts: binary image data, data type semantics for managing modular image analysis, and image metadata definitions for recording contextual information. Image data in OME are stored as time-lapse, 3D, multispectral files ("5D images") (6, 7). Data type semantics for OME are designed to allow analytic modules to be strung together in a flexible and simple fashion and are described in detail below. Image metadata describe the optics of the microscope, the filter sets, the objective lens, etc. We hope that microscope manufacturers will agree (through OME or other projects) on a common format for metadata describing microscope hardware and image acquisition. Image metadata also describe the experimental setup, including genes under study, fluorophores, etc. Whenever possible, OME metadata definitions derive from preexisting ontologies such as

the Medical Subject Headings (MeSH) and the Microarray Gene Expression Data Society (MGED) and those being developed by the Minimal Information About a Microarray Experiment (MIAME) effort (8).

OME is designed to connect a desktop computer to an Oracle or PostgreSQL relational database using a standard client-server paradigm (Fig. 2; blue cylinder). The relational structure of

oms into a single database representation. The OME file format is used when image data and metadata must be translated into a file for transport between OME databases or for storage outside of a database. In OME files, each piece of data is associated with a tag (such as <filter_wavelength>) that defines its meaning in extensible markup language (XML), providing a vendor-neutral file format that conforms to public

Web-compliant standards (Fig. 2; green box). We anticipate that commercial software tools will eventually be able to interact with the OME database as clients or possibly even directly read and write OME files.

Data Semantics for Image Analysis

Searchable image archives are useful, but we really require informatic systems that can extract and store quantitative information derived from images. Typically, image analysis involves several processing steps, but the precise steps and their sequence necessarily depend on the properties of the image and on scientific goals. We therefore require an extensible toolbox of algorithms (including fourth-generation languages such as Matlab) that can be applied in different combinations to different images. Consider the

problem of tracking labeled vesicles in a time-lapse movie. A segmentation algorithm finds the vesicles and produces a list of centroids, volumes, signal intensities, etc.; a tracker then defines trajectories by linking centroids at different time points according to a predetermined set of rules; and finally a viewer displays the analytic results overlaid on the original movie. As designers of OME, we cannot anticipate exactly which tools work best for vesicle tracking. Instead, we must build general mechanisms for linking an analysis toolbox to images and storing analysis results.

Although we typically think of databases as storage systems, databases are also an ideal mechanism for linking independent pieces of software together in a modular fashion. Conceptually, the data path in OME is from one analytical module to the next (Fig. 2; dashed box), but

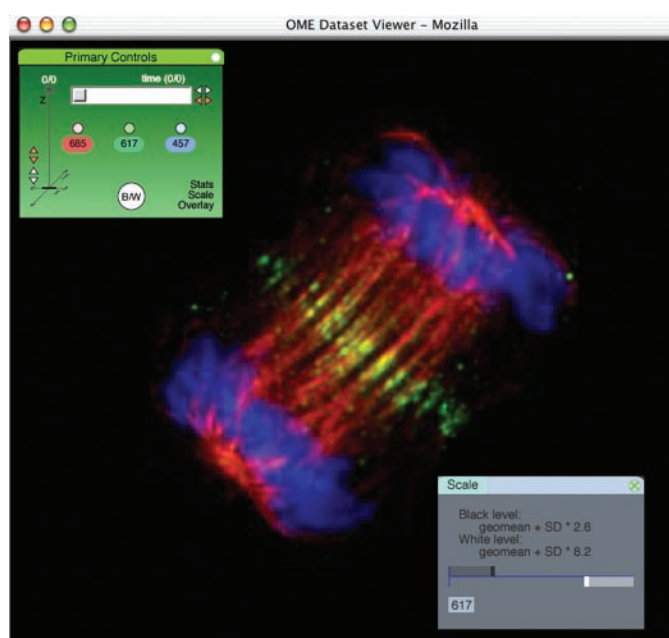


Fig. 1. Applications for quantitative imaging. The image shows an Xlk2 cell during the process of cytokinesis stained for DNA (blue), microtubules (green), and the aurora-B protein kinase (red) (13). Although the image demonstrates the relative localization of different cellular components and structures, quantitative analysis reveals specific characteristics that can be used to assay the effects of inhibitors or expressed proteins. For example, integrating the signal from a DNA-specific fluorophore might reveal defects in chromosome segregation during mitosis. Measuring the overlap of microtubules and aurora-B [for example, using a cross-correlation analysis (14)] within a subregion of a dividing cell might be used to assess effectors of cytokinesis. The image is displayed within the Image Viewer that ships with OME. The Viewer includes support for displaying multi-dimensional image data (top left) and some of the associated metadata about each image (bottom right).

OME makes it easy to access images on the basis of content and meaning: "Find all images of HeLa cells recorded by Jason in 2002." Queries of this type are accomplished via an application layer comprising import and export routines, interfaces for analytic and visualization tools, and ancillary software. As an aid to performance, binary image data are stored in a file system (a repository) accessible only to OME (Fig. 2; red box). Images from commercial file formats are imported into OME using a translator that reads the image data and converts them into a multi-dimensional image repository format. Any metadata stored with the image (usually in a "header" that precedes the pixel data) are extracted from an input file and stored in the appropriate database tables (Fig. 2; blue table). The net result is the conversion of a polyglot of commercial file for-

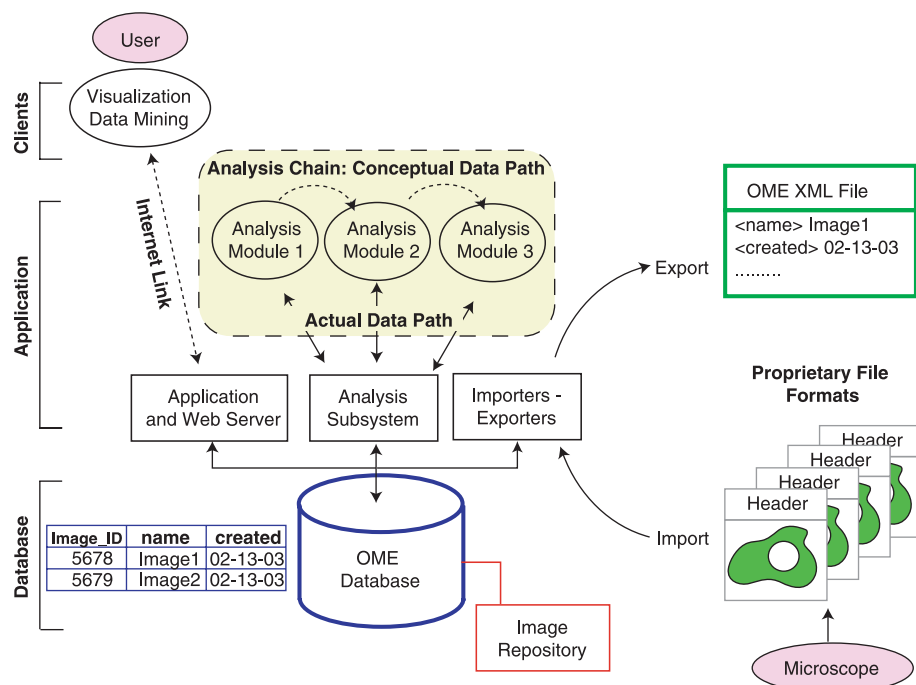


Fig. 2. The database is the interface: OME architecture. OME is constructed as a standard three-tier application with a relational database that stores information in a table-based structure (blue table), an application server that processes data, and a client that lives on the user's desktop and communicates via the Internet (that is, via IP). Multiple clients can communicate with OME, including Web browsers, commercial microscopy software, and data-mining applications. The OME data model is instantiated via a relational database ("OME database"; blue) in which metadata are stored in tables as specified by the schema, and binary image data are stored in a trusted file system (the "image repository"; red). When data are transported between databases or stored in a flat file, metadata and image data in the database are translated into XML ("OME XML File"; green). The OME database communicates with analysis modules via a subsystem ("analysis subsystem") that ensures the consistent treatment of semantic data types. The analysis modules also calculate and store the history of the analysis chain (see supporting online material). When analysis modules are chained together, each communicates independently with the database ("actual data path"; yellow block), even though the conceptual path is from one module to the next ("conceptual data path"). Existing commercial or independent software tools can read OME data without substantial modification. OME itself is open-source and available through a Lesser General Public License, but applications that talk to it can be either open or proprietary.

in practice, each module communicates independently with the database. The advantage of this architecture is that the problem of building links between analysis modules written in different computer languages is simplified to the task of linking each module independently to the database using known methods. For this to work, however, the output of one module must match the input of the next module. The OME data model therefore includes a set of semantic data types that describe analytic results such as "centroid," "trajectory," "maximum signal," etc. Semantic typing defines the types of relationships that a data type can participate in, and thus determines which analytic modules can use the data as inputs and outputs. However, the process of data analysis is tightly tied to prior knowledge of the biological system, the experiment, and the properties of the analytic routines. It is simply not possible to create a standards body that will rule on which definitions of "centroid" are valid and which are not. However, a database can solve this problem by linking each result to an operational record of the data-processing steps that produced it, including the algorithm used

and the states of any settings or variables. Thus, semantic data types such as "centroid" can be defined broadly and then given specific meaning by the recorded history of their derivation. In this way, we can judge each result in light of the methods that generated it and determine the accuracy of measurements a posteriori, given the known operation of the analytic algorithms and characteristics of the data.

A final challenge for OME is providing a mechanism for adding new analysis modules. In some cases, the inputs and outputs of the new module correspond to existing OME semantic data types (supporting online text). A more complex situation arises if OME lacks the necessary data types for a new module to interact with the system. In this case, the database must be augmented with tables to store the new data and present the changes to the user interface. This is a challenging problem in database design and represents a type of extensibility that is absolutely critical, but usually absent, in most bioinformatics software. Our solution is to specify the data requirements (the inputs and outputs) of an analytical module in an XML description written

to the OME XML specification. OME is designed to then create the necessary tables on the fly. The net result is an analytic system that is extensible, modular, and language-independent.

Image Informatics in Practice

The OME system is being developed as an open-source collaboration between academic labs and commercial hardware and software imaging companies (9). OMEv1.0 (10) demonstrated the utility of general-purpose image informatics software (11). A tutorial demonstrating this system is available (12). A system with features necessary for general use is being developed as OMEv2.0 for release late in 2003 (10). Our software is open-source, but commercial code plays a vital role in modern digital microscopy and image analysis. OME is therefore designed to integrate effectively with commercial code. Like DNA and protein sequence analysis, biological imaging must develop features of an information science to meet the demands of screening and hypothesis-driven analysis. Macromolecular and imaging bioinformatics have many things in common, but the complexity and unstructured nature of biological images present a unique set of challenges in data analysis and interpretation. We are confident that if commercial and academic microscopists can solve the informatic problems that currently make quantitative analysis of microscope images difficult, quantitative image analysis will assume an important position in the future of bioinformatics.

References and Notes

1. C. M. Waterman-Storer, G. Danuser, *Curr. Biol.* **12**, R633 (2002).
2. T. U. Mayer *et al.*, *Science* **286**, 971 (1999).
3. P. Gonczy *et al.*, *Nature* **408**, 331 (2000).
4. K. Huang, J. Lin, J. A. Gajnak, R. F. Murphy, *Proc. IEEE Symp. Biomed. Imaging* **2002**, 325 (2002).
5. See www.openmicroscopy.org.
6. H. Chen, J. W. Sedat, D. A. Agard, in *Handbook of Biological Confocal Microscopy*, J. B. Pawley, Ed. (Plenum, New York, 1989), pp. 141–150.
7. P. D. Andrews, I. S. Harper, J. R. Swedlow, *Traffic* **3**, 29 (2002).
8. A. Brazma *et al.*, *Adv. Biochem. Eng. Biotechnol.* **77**, 113 (2002).
9. See www.openmicroscopy.org/participants2.htm.
10. See www.openmicroscopy.org/technology6.htm.
11. M. Platani, I. Goldberg, A. I. Lamond, J. R. Swedlow, *Nature Cell Biol.* **4**, 502 (2002).
12. See www.openmicroscopy.org/OMEV1/OME_V1.html.
13. R. R. Adams, M. Carmena, W. C. Earnshaw, *Trends Cell Biol.* **11**, 49 (2001).
14. R. N. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1986).
15. We gratefully acknowledge helpful discussions with our academic and commercial partners (9); G. Danuser; and members of the Swedlow, Goldberg, and Sorger groups. Research in the authors' laboratories is supported by grants from the Wellcome Trust and Cancer Research UK (J.R.S.), and NIH and the Harvard Institute of Chemistry and Cell Biology (P.K.S. and I.G.). J.R.S. is a Wellcome Trust Senior Research Fellow.

Supporting Online Material

www.sciencemag.org/cgi/content/full/300/5616/100/DC1
SOM Text
Fig. S1
References