BARBARA FOSTER

Informatics for the Scientific Lab:

An Enterprise Image-Analysis Revolution?

There's new science on the horizon. It's built on Informatics, the ultimate relational database, and it's coming to a lab near you! Combining images, text, numerical data, graphics, and the power of the Internet, image-enabled informatics is one of the fortunate offsprings of the Human Genome project, and its more commercial sibling, high-throughput drug screening (HTS). Both efforts are generating information at a prodigious rate, which all needs to be cross-correlated and totally searchable on a global scale.

Now, Informatic programming is appearing in academic, research, and commercial facilities. And it's clearly not just for biomedical fields; it has already spread to materials, and its use in semiconductors is both logical and imminent. The technology looks like a boon for the user and an opportunity for providers of image-analysis tools.

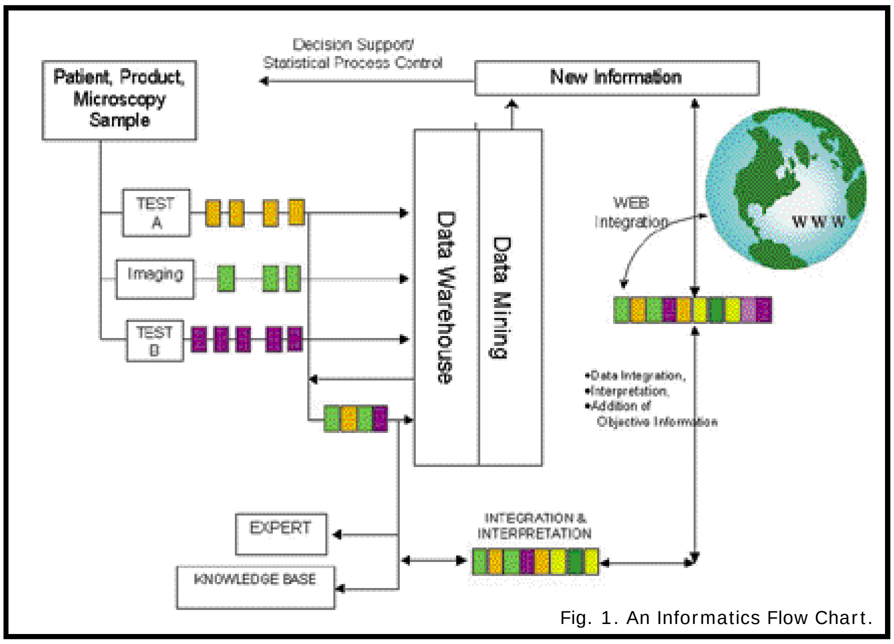

Fig. 1. An Informatics Flow Chart.

MICROSCOPY: PRESENT AT THE CREATION

Simple image databases are, of course, commonplace in microscopy. The best ones allow microscopists to register and then search on everything—magnification, numerical aperture, contrast technique, sample number, type, and method of preparation, etc. Informatics goes much further. Silently and without much notice, all forms of biologically-oriented microscopy have already moved from that first-level platform to become embedded in the new Informatics.

In an exotic technology twist, confocal microscopy, still used primarily in research, has given birth to a new scanning technology for "gene chips"—arrays of up to 10,000 micro-dots of diagnostic chemicals or biological components printed onto microscope slides via either conventional inkjet or photo-lithography-type technologies.

Used both for HTS and in innovative genetic screening for disease, the resulting fluorescent patterns require the high sensitivity and spatial resolution typical that confocal microscopy provides. General Scanning (Watertown, MA) announced one of the earliest confocal-type scanners; Genetic Microsystems and Virtek (both from Woburn, MA) quickly followed, as did confocal veteran A. E. Dixon (Waterloo, ON).

All of the systems use third-party software to convey the scanned data into an organized Informatics matrix. Informatics integrates the massive amount of data resulting from the scan in terms of position in the matrix, pattern and intensity of fluorescence in each spot, and then connects that information with the corresponding gene or protein fragment.

While this does not seem to be an overwhelming problem at first glance, the gene and protein libraries are vast, and different methods and search patterns are used for elucidating and finding the fragments. See eBioinformatics' Web site (*www.ebioinformatics.com*) and take a short tour through their BioNavigator for quick insight into this problem. BioNavigator provides a very simple user interface and easy-to-follow protocol which masks the complexity behind the scenes. Other companies providing gene-oriented informatics services include InforMax, Bio-Discovery, GeneCore and Double-Twist (previously Pangea).

Pathologists, usually conservative and slow to embrace new technology, are actually leading the Informatics pack with more conventional microscopies. Electronic medical records (EMRs) are beginning to include SEM, light, and confocal images along with the more routine X-rays, blood tests, gross anatomy images and doctors' comments.

Ironically, while genomics and HTS gave birth to Informatics, its spread is now being pushed by new consumerism in the medical industry. Information is big business—the ultimate goals are to provide:

• Patients with full access over the Internet (on a secure channel) to his/her medical records at any point during diagnosis and treatment.

• Physicians and researchers with worldwide availability of diverse forms of information which can be used to improve diagnoses and speed investigation.

• Students with unlimited resources for distance learning.

• Hospitals, HMOs and other business bodies with one-stop recordkeeping, which integrates EMRs with the local lab information system (LIS) and administrative functions such as patient tracking, scheduling, and billing and ordering.

• Finally, full but secure availability through a LAN or on the Web.

Applying this model to the analytical world at large means that Informatics will have a profound effect on all human knowledge, from how and what we learn, to how we investigate, to how we commercialize, produce, and analyze products.

AN ASTOUNDING PICTURE

Many prototypes for informatics are coming out of the Division of Pathology Informatics at the University of Pittsburgh. DPI's Director, Dr. Michael Becich, speaks and writes energetically on the topic and paints nothing less than an astounding picture of the power of Informatics in both integrated reporting (which he refers to as "Information Therapy") and data mining.

Fig. 1, at left, illustrates the relationship between data influx and outgo in the image-enabled Informatics universe. All information starts with the patient, product, or microscopy sample (upper right corner). In a more generic microscopy application, all information would start with the part under analysis.

When a sample is brought to a microscopist, we ask the engineer or technician to tell us about its background: Where has it been? How has it been treated? What behavior leads you to believe that there is a problem? Sound like a visit to the doctor's office? Then, just as a doctor conducts anatomical (ex., Test A) and clinical (ex., Test B) pathology tests, supported by a variety of images (X-rays to micrographs), a microscopist would perform the necessary physical and chemical testing, again supported by images. This process gives rise to the "structured data" seen in the middle of the diagram.

How the information is handled at this point and where it goes from here is the crux of Informatics. In the simplest case, the information will go into an integrated report. Microscopists have been doing this step manually for generations. More recently, products such as Media Cybernetics' Image-Pro Plus have made the job easier and more automated. Dr. Becich's concept is shown in Figure 2 (pg. 39) and truly encapsulates the concept of Informatics.

Note that this is a Web-based report which can be accessed 24 hours a day, 7 days a week. It can be customized to the patient, the doctor, or for reference by an outside source to protect confidentiality. Secondly, it integrates a variety of tests (clicking on any of the older test dates would activate a window with that particular data), images, and the pathologist's report, as well as key demographics to help the patient understand his particular situation vis-à-vis the population at this specific hospital. There are also useful URLs, should he decide to research his condition further.

Ideally, the program which constructed this report would have automatically searched the Web for those addresses. And, for answers to further questions, the pathologist has listed his phone and pager numbers, as well as a hot link to his e-mail.

Continuing along the architectural diagram through the lower middle branches shows information's traditional paths: most typically to storage, or, in the case of a university or teaching hospital, into a knowledge base to be used for training.

When information is sent to a Data Warehouse where massive amounts of data can be organized for "mining" or integrated and mixed with other information to form new bodies of knowledge, Informatics takes over. At this stage, for example, data from a laboratory information system (LIS) may combine with scientific research on the manifestation of a specific disease and population statistics from a globally remote portion of the Internet to create a new model for disease propagation. The resulting whole is much larger than the sum of its parts.

CHALLENGES: TECHNOLOGY AND TERMINOLOGY

As with any computer-based discipline, the first challenge will be to learn the latest lingo. For example, "sample number" may be replaced with "accession number". Then there are the abbreviations. Informatics falls under the purview of information technology (IT),

Indicate #020 on Reader Service Card

Indicate #021 on Reader Service Card

so, in the world of Informatics, many institutions may get to be known by only by their initials!

For example, one recent paper cited lead-poisoning statistics collected by the CCF, as part of the OCLPP, in accordance with the CDC and ODH, with final output as an HMR, merged with a database connected by OLE DB drivers. Translation: the paper refers to work done by the Cleveland Clinic Foundation laboratory, collected as part of the Ohio Childhood Lead Poisoning Prevention Plan in accordance with recommendations from the Centers for Disease Control and Ohio Department of Health. The output form was a Heavy Metal Report in which data were merged with a database connected by Object Linking and Embedding DataBase drivers.

But there are more serious technical challenges. First and foremost in the mind of microscopists is image quality. High-resolution color images of diagnostic or analytical quality equate to huge files. How do you transport them in the Informatics model? Or do you transport them? Are they better left on a server, with access via the Internet?

Secondly, how do we integrate new information with existing databases or laboratory or hospital information systems (LIS and HIS)? Known as "legacy systems", these dedicated information superstructures may be the current archive for tens of thousands of images and their related information. How do we get them talking to one another? Some sources are looking to Microsoft's ubiquitous ACCESS, but is it powerful enough?

The data management challenge in Informatics is so huge that headhunters in the arenas of genomics and protenomics (the study of disease sources derived from proteins that make up genomics' genes) have been chartered with finding high-level computer and database skills first. The employer will provide training in the necessary science!

Finally, there is the Internet. Connectivity, you've probably noticed, is not always what it should be. A given institution may have a LAN connecting all its computers in-house, or may even be connected over larger geographies via WAN, but all LANs and WANs have built-in firewalls to block access to hackers and competitors. The latest wrinkle for telepathology is dealing with the difficulty accessing images through those firewalls.

Conversely, if all this information is floating around the Web, can we simultaneously grant full access to researchers and clinicians while protecting the privacy of the patient or, in the industrial realm, the product?

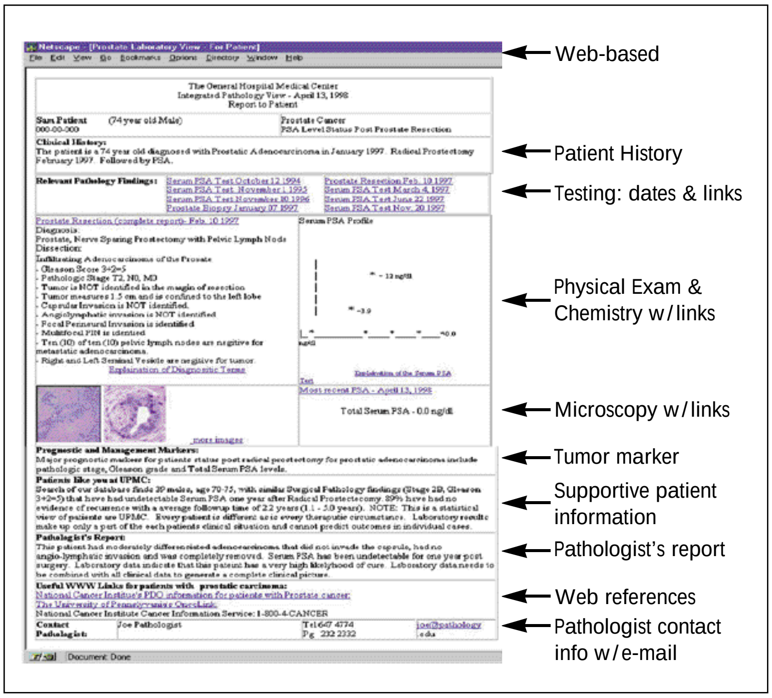

Fig. 2. An integrated, Web-based report with integrated information from chemical tests, images, text, and numerous active links to previous tests, pertinent URLs for further research, and the doctor.

GLIMPSES FROM CONFERENCE

This past October, DPI hosted the fourth annual conference on Advanced Pathology Informatics, Imaging, and the Internet (APIII). Their program reveals the reach of Informatics: from distance learning to research and on to clinical practice. Presentation and tutorials addressed key issues including data annotation (frameworks for captures and formation data for easy integration into Informatic systems), ways to integrate legacy systems, how to use digital technology to become more cost-effective, and how to convert pathologists into information managers. These papers are a wonderful resource for those of us from other disciplines! (See *http://apiii.upmc.edu/exh.htm*.)

Whether you are the director of a polymer lab, a failure analyst on a semiconductor fab line, a biological researcher, or a potential new provider of these tools, if Informatics is in your future, a review of the abstracts from this meeting would help you build a strong foundation quickly without having to reinvent the wheel.

FROM CRADLE TO GRAVE

It's no wonder that Informatics is spreading like wildfire. Inherently, it grants us the ability to track complex processes and to communicate findings and expand knowledge globally. Bioinformatics and pathology-based informatics are facing the early challenges and building the foundations.

The step to a parallel industrial universe is small—Informatics will enable us to integrate physical and chemical testing, imaging, and process control, from R&D to commercialization, from the classroom to quality control and failure analysis and back to the production line. Microscopy is providing (and will continue to provide) images which are at the core of much of the Informatics revolution, creating a powerful gateway to a new global world of information. ◆