

## ◆ *Bioinformatics and Publishing*

Speakers:

**Johanna McEntyre**  
National Center for Biotechnology  
Information  
Bethesda, Maryland

**Amy Brand**  
CrossRef  
Burlington, Massachusetts

Reporter:

**Marilyn T Regan**  
Stroke Basic Science Editorial Office  
Massachusetts General Hospital  
Boston, Massachusetts

Bioinformatics is an approach to the biologic sciences that combines biology, computer science, and information technology. Life science is no longer purely a laboratory science; it is an information science, too.

Johanna McEntyre discussed the public nucleotide-sequence database, GenBank, the US partner in an international collaboration with databases in Japan (DDJB) and Europe (EMBL). Data are exchanged daily, so content in all three is the same. However, bioinformatics includes other data types, such as protein structures, whole genome sequences, single nucleotide polymorphisms (snps), and bibliographic research articles and abstracts (PubMed). Bioinformatics aims to provide scientists with useful tools and interfaces with which they can explore the available information.

One approach is to create better links between the scientific literature and the molecular biology information. McEntyre said the Bookshelf Project, which provides an electronic collection of books, works in collaboration with book publishers and links books in a useful manner to public databases, such as GenBank and PubMed. After an initial pilot project with the third edition of *Molecular Biology of the Cell* (by Alberts and

others, published by Garland), the Bookshelf now has 10 books available on topics from basic biology to cancer to retroviruses. The books are linked to PubMed abstracts by keywords, and a book's entire content can be searched directly.

The books, kept in a database, are

*Bioinformatics is an approach to the biologic sciences that combines biology, computer science, and information technology. Life science is no longer purely a laboratory science; it is an information science, too.*

in XML. When someone searches the Bookshelf and locates a page of interest, the page is converted to HTML with an XSLT processor. Books are viewed with an Internet browser one section at a time. The amount of navigation within a book depends on the policy of the publisher; most allow full access and browsing via a table of contents and search access.

Future development of the Bookshelf will include deeper links to various public databases to provide richness to the books and to provide background information for users of the biology databases.

Amy Brand then discussed a cross-publisher citation-linking platform called CrossRef. CrossRef does not aggregate full

text or sell or provide information. It is multidisciplinary and links journal articles in bioscience and medicine. The linking network includes books and conference proceedings.

CrossRef uses a digital object identifier (DOI) link, an alphanumeric string that allows one-click access at the article level, regardless of the publisher or journal. It enables a researcher to navigate the literature without having to switch to a search interface to find a cited article at another publisher's Web site. Publishers can update their online content in one place and avoid the problem of links that no longer work. Furthermore, they do not have to sign numerous bilateral linking agreements with other publishers.

Like Bookshelf, CrossRef is free to users, such as publishers, agents, libraries, and affiliates. Users are required to deposit a limited amount of metadata. Members are active, however, and register new content daily. Fees are minimal and are paid by the publisher, although there is no cost for publisher-to-publisher links.

Bioinformatics provides a means to finding the most newsworthy discoveries daily free of charge. The information will improve not only the rate at which new scientific information is shared but also the rate of discovery itself. It will require new safeguards to ensure that the information provided to the public and scientists is accurate.

The challenge for publishers will be to create a system, such as peer review, that provides rigorous rapid review of new scientific information. Bioinformatics has the potential to revolutionize the dissemination of scientific discovery in much the same way that the modern computer revolutionized communication. The goal of publishers will be to provide a system that adequately accommodates the explosion of scientific data in a responsible manner. 