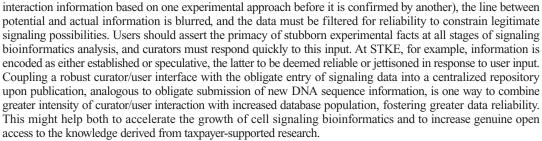# A Two-Way Bioinformatic Street

The rapid emergence of Web-based bioinformatics systems reflects the research community's attempts to embrace the biological complexity uncovered by high-throughput genome, transcriptome, and proteome data acquisition and the sheer size of the modern scientific endeavor. If information systems can match this complexity, biology will be enriched as a result. If not, scientific excitement may paradoxically be dampened by data flow. The question is, how should biological information systems and the relationship between those who use them and contribute to them further evolve?

Before the advent of high-throughput research genres such as genomics and proteomics, fields already replete with information such as cell signaling (focused on uncovering the flow of information through a cell) advanced through scientists cross-communicating and assembling and synthesizing their own information. Because deciphering cell signal transduction is crucial to understanding normal and diseased biological processes, curating reliable data in the field has become at once a necessity and an enormous challenge, given the massive increase in available data. Cross-communication between the users and curators (also enlisted as experts, authorities, and gurus) of databases is now at the heart of enhancing data reliability. Efforts including the Connections Maps at *Science*'s Signal Transduction Knowledge Environment (STKE) and pathway-building at Biocarta, Inc., exemplify Web-based databases that include an avenue for making the curator/user interface a two-way street. Enhancing curator/user exchanges might make visiting these environments a more lively and entertaining experience and increase their usage, large-scale participation being the sine qua non of usefulness to the scientific community.

A primary ingredient for massive exchange of information among multiple bioinformatics tools and databases is curator tagging of input information to enable proofreading and data correction. Minor changes in a protein or DNA sequence entered into a gene or protein database can be corrected and generally will not propagate error throughout the entire informational system. Bad information in a protein interaction or pathways database is trickier. If information gatherers skip a step (for example, entering interaction information based on one experimental approach before it is confirmed by another), the line between potential and actual information is blurred, and the data must be filtered for reliability to constrain legitimate signaling possibilities. Users should assert the primacy of stubborn experimental facts at all stages of signaling bioinformatics analysis, and curators must respond quickly to this input. At STKE, for example, information is encoded as either established or speculative, the latter to be deemed reliable or jettisoned in response to user input. Coupling a robust curator/user interface with the obligate entry of signaling data into a centralized repository upon publication, analogous to obligate submission of new DNA sequence information, is one way to combine greater intensity of curator/user interaction with increased database population, fostering greater data reliability. This might help both to accelerate the growth of cell signaling bioinformatics and to increase genuine open access to the knowledge derived from taxpayer-supported research.

Another critical element in developing cell signaling databases is providing access to the raw data for swapping among various software platforms for visualization and analysis of biological information, including cell signaling pathways. Molecular interaction data from the Biomolecular Interaction Network Database (BIND), for instance, can be exported to an assembly-based information software system such as Cytoscape, greatly enhancing the value of the underlying data set. The availability of curator-tagged input data wrapped for portability should promote efficient distribution of data entered at any port, into the entire network of signaling tools. It will also improve curation, avoid duplication of effort, and eliminate tools that lack content for application. The gurus should argue strongly for it.

Used intensively, a well-connected array of bioinformatic tools can form a computational "working memory" for assembling biological information from specialized organism, cell system, and molecular data that the scientist can access for designing new experiments that are maximally informative. Movement toward centralized electronic pathway submission and improved data portability will make it possible to integrate new sources of data, including cellular locations of signaling complexes and components, quantitative aspects of signaling, and pharmacological data, into current pathway analysis databases and tools. This should be a strong motivation for the scientific community to increase its collective investment in the next phase of signal transduction bioinformatics development.

Lee E. Eiden

Lee E. Eiden is chief of the Section on Molecular Neuroscience and chairs the Bioinformatics Users Group in the National Institute of Mental Health Intramural Research Program, Bethesda, MD.