The Human Brain Program (HBP) organized by the Office of Neuroinformatics of the National Institute of Mental Health, National Institutes of Health, is a broadly based federal research initiative sponsored, in a coordinated fashion, by 15 federal organizations from four federal agencies (http://www.nimh.nih.gov/neuroinformatics/index.cfm). The scientific goals of this initiative are to enable the progress of neuroscience research through the creation of a Web-based set of distributed and federated databases, analytical and modeling tools, and simulators.

Currently, neuroscientists collect complex data in ever-increasing amounts, fostering increased specialization, with resultant challenges to integrate data between and across levels of interaction, control, and function. The sheer quantity and complexity of the data are such that the field of neuroscience would benefit considerably from an information management system for its experimental data. The field should enhance its wealth of ever-increasing empirical data, accumulated from its many disciplines and experimental approaches, by developing appropriate databases and a greater capability for both theory development and simulation models.

The focus section in this issue of JAMIA, on HBP research progress in neuroinformatics, collects three representative works of HBP grantees on the tools and methodology to support the bioinformatics aspects of neuroscience research.

The amount of neuroscience information increases significantly as the HBP and related clinical research move forward. The paper by Daniel Gardner et al. addresses one critical issue in interoperability, i.e., facilitate neuroscience research and information exchange through a common data model (CDM), which is generalized from two prototype neurophysiology databases. These authors apply the emerging extensible markup language (XML) standard to implement data exchange between any CDM-derived data models. The authors state nine design goals that meet the essential criteria of a good data
model, and adopt the data-driven design philosophy, which, they explain, is well suited to brain information. The CDM has five root classes, or “super classes”—data, site, method, model, and reference—that are a reasonable abstraction of the problem domain. The authors believe that when fully implemented using biophysical description markup language (BDML), as well as the hierarchic attribute value implementation of controlled vocabulary, the proposed CDM will have the capacity to mediate among disparate neuroscience database projects and similar resources with compatible data and data models.

The experimental work of the Yale research group on neuroscience data analysis provides yet another perspective on data modeling and knowledge representation. The traditional paradigm of reductionism works reasonably well in physical sciences, but the biological domain requires a different research paradigm. A major goal for neuroscientists is to have a sound theoretic foundation that can cut across multiple biological levels, from the genetic level up through the synaptic, neuronal, network, and brain pathway levels, ultimately reaching the behavioral level. The ability to capture, relate, and analyze diverse types of data at multiple levels of abstraction is one of the central challenges of neuroinformatics.

The work reported by Perry Miller et al.3 on the integrated data analysis of a particular model system of neurobiology, i.e., the olfactory system, explores new paradigms of biological research to study new relations among findings obtained at individual levels of representation. Their research paper provides an overview of SenseLab; in particular, the introduction of a flexible data model called EAV/CR (entity-attribute-value with classes and relationships) to integrate olfactory and associated data of four different databases for neuroscience research and clinical studies later on. The modeling overhead of EAV/CR, however, may not be suitable to model large amounts of very homogeneous data that have been mass-produced by high-throughput instruments, such as microarray analyzers. Although the current focus of SenseLab’s activities is on basic neuroscience and neuroinformatics research, the work has a range of potential clinical and other real-world applications.

We hope that the ideas and results reported in this focus section will suggest new and better ways to develop tools to support bioinformatics research and will lead us to the next generation of bioinformatics tools and management systems. Another HBP focus section, on the progress of enabling technology for brain imaging aspects of neuroscience research, will appear in a future issue of the Journal.—STEPHEN T.C. WONG, STEPHEN H. KOSLOW

References


Affiliation of the author: University of California, San Francisco, California (STCW) and National Institute of Mental Health, Bethesda, Maryland (SHK).

Correspondence and reprints: Stephen T.C. Wong, PhD, PE, UCSF, Department of Radiology, Box 0628, UCSF 505, Parnassus Avenue, San Francisco, CA 94131; e-mail: <swong@radiology.ucsf.edu>.

Received for publication: 9/27/00; accepted for publication: 9/28/00.