LNI Laboratory of Neuroinformatics
Weill Medical College of Cornell University

The Laboratory of Neuroinformatics

Department of Physiology and Biophysics
Weill Medical College of Cornell University

Daniel Gardner, Ph.D.
Principal Investigator

THE MISSION OF THE LABORATORY OF NEUROINFORMATICS

Contemporary neuroscience acquires and analyzes brain data of unparalleled magnitude and diversity. Extracting information and insight from these data requires the development of new and creative methods and syntheses, many of which derive from advances in computation and in networking. In the Laboratory of Neuroinformatics, our work is designed to develop standards for data exchange and interoperability among resource databases that span contemporary brain and informational sciences. The goal is the development of a new discipline—neuroinformatics—poised to enhance neuroscience as genomic bioinformatics has enhanced genetics. Our principal funding is provided by the Human Brain Project and the National Institutes of Health [NIH] via NIMH and NINDS, with past support from the National Science Foundation [NSF], NINDS, and NIMH. NIH's new data sharing mandate makes the mission of the Laboratory of Neuroinformatics increasingly significant, and our work will benefit from additional funding developed from government and corporate sources.

JOIN US

We welcome applications from neuroscientists with a background in, and understanding of, bioinformatics, computational methods, computer science, medical informatics, or information theory. Send a CV and the names of three references to: dan@med.cornell.edu.

BACKGROUND

Biomedicine is increasingly an information-rich science. Information-driven contemporary research explores:

- The nature and coding of biological information transmitted and processed by the brain
- The relationship between information processing of cells and organisms
- Genetic, molecular, somatic, and phenotypic consequences of information processing
- Errors of information processing and their relation to disease

Neuroinformatics is the integration of brain information from gene to behavior. It merges neuroscience, information theory, computer science, and a broad range of other disciplines.

The Human Brain Project supports research and development of advanced technologies through cooperative efforts among neuroscientists and information scientists. A coordinated effort of fifteen federal institutes across four federal agencies (NIH/NSF/NASA/DOE), it currently funds 36 projects, including our research. A major goal is a federated World Wide Web based information management system for neuroscience, linking interoperable databases and data management tools to provide a better understanding of the structure and function of the brain.

The Laboratory of Neuroinformatics [LNI] in the Department of Physiology and Biophysics at Weill Medical College of Cornell University is dedicated to research in neuroinformatics, both informational and computational. Our methods meld neuroscience with advanced technologies of computer and information sciences. The following pages present our goals, detail our current and projected areas of exploration and development, and offer contact information.

Our Goals

Our multiple projects, together with the work of collaborators within and beyond the Human Brain project, are designed to enable neuroscientist users to:

- archive, exchange, and re-analyze data from remote labs along with data from local sensors
- enable and test models of neural function both locally and remotely
- design and test hypotheses about neural coding
- formulate hypotheses similarly informed by a network of related data and hypotheses
- relate multiple levels of brain function—and beyond brain function to genetic and genomic and other physiologic data as well.

Research in the Laboratory of Neuroinformatics

I: Current and ongoing project phases

- ***Neurophysiology Database Development***

[This thrust is currently funded by the LNI's main Human Brain Project grant, with additional past support from NSF, NINDS, and NIMH]

Sensor-based brain data including time series neurophysiological records (spike train, patch clamp, EEG, MEG) convey and report information processing in the brain. The initial thrust of the scientific program of the LNI designs and supports archives and analyses of such records. The goal is to provide insights into fundamental properties of neuronal ensembles in the same manner as archives of protein and nucleic acid sequences have revolutionized molecular biology.

Towards this end, we are developing databases to collect and archive neurophysiological recordings from brain neurons and regions. These databases, resembling yet transcending sequence archives, are designed to exchange and deliver actual brain recordings world wide, permitting further analysis and comparison to other local or global data. The goal is an understanding of the neural code, that pattern of activity in each neuron of the brain that encodes and transmits sensation, motor commands, cognition, and memory.

Neuroscience data are richer and more complex than sequence data, which can be searched and expressed as simple text strings. Neuroscience data are instead formatted as, for example, two- or three-dimensional images, as arrays of digitized sensor data such as electrode current or voltage, or as histograms of signal amplitudes or durations. To support these, we have developed new methods and data structures that are intuitive and easy to implement. New computer languages such as Java and XML allow the construction of data archives that can be accessed with a web browser from any contemporary computer—Mac, PC, Unix, Linux, etc. These languages provide state of the art tools and methods for data entry, storage, visualization and search, each designed to aid investigator-directed retrieval and analysis of information. Browsing or directed queries of textual metadata use our Hierarchic-Attribute-Value schema that permits both broad and targeted searches of the database contents. Data are cached transiently, and displayed, together with identifying metadata, using our Java Virtual Oscilloscope. An extensive Java midlayer provides public methods for linking client and database while shielding users from internals.

This neurodatabase system allows users to download the actual datasets that comprise information displayed in journal figures, so that they can be viewed on expanded or contracted time scales, and analyzed with algorithms other than those used in the original study. This adds value to data acquired in individual laboratories, enlarging the total population analyzed.

• ***Towards Interoperable Neuroinformatics***

[THIS THRUST IS ALSO CURRENTLY FUNDED BY THE LNI'S MAIN HUMAN BRAIN PROJECT GRANT]

A second, newer, project is directed toward developing open standards and techniques to coordinate and interchange disparate data types between different Human Brain Project resources. Each resource adopts a data model, syntax, and semantics that covers a specific domain of interest, and these domains span multiple levels, questions, techniques, and preparations. HBP resources include libraries of fMRI images and MEG recordings, dendritic tree morphologies, three-dimensional EM reconstructions of spine arrangements on dendrites, conductance properties of ion channels underlying synaptic currents, classes of olfactory receptors, spike trains of cortical neurons, atlases of neurotransmitters in mouse brains, and flat maps of the cerebral cortices of humans and monkeys. Interoperability between such resources requires agreement on technical and lexical standards to describe data, manage queries, design tools, models, and hypotheses, and allow them to be referenced, exchanged, transferred, or converted.

Toward this end, we are designing and codifying a scientific markup language called BrainML to serve as a functional standard underlying interoperability between resources. As no single such model can serve such a wide field, or accommodate evolution of techniques and data types, BrainML is being structured as a suite.

The BrainML metalanguage BrainMetaL provides an underlying layer of structure and abstract semantics as a foundation for characterizing experimental science. BrainMetaL includes abstract classes for:
• data, entities, methods, models, and references,
• standard scientific quantities, units, and prefixes, and
• hierarchic lexical structures to organize controlled vocabularies of descriptors.
Such abstractions are sufficiently basic to accommodate physical, chemical, psychological, and clinical correlates of biophysics or neuroscience.

This architecture is designed to enable broad applicability for BrainML. Biomedical resources can utilize our BrainMetaL language and technique development efforts to design specialized yet compatible data models, libraries, and methods for interoperability. Brain ML will thus be expanded to encompass other resources being developed in neuroscience, biophysics, and physiology.

• ***The Laboratory of Neuroinformatics and the Human Brain Project***

[AS NOTED ABOVE, MOST OF THE CURRENT EXTRAMURAL FUNDING OF THE LABORATORY OF NEUROINFORMATICS COMES FROM THE HUMAN BRAIN PROJECT, VIA NIMH AND NINDS, WITH PRIOR SUPPORT FROM THESE INSTITUTES AND THE NSF.]

The LNI has been an effective developer and advocate for the application of cutting-edge information technology within the Human Brain Project. Towards interoperability among projects, we have designed, promoted, and provided proof-of-concept for:
• Use of XML, the emerging-standard extensible markup language, in neuroscience
• Layered architecture for scientific markup languages
• Java for multiplatform delivery and technology evolution
• Archiving and delivery of actual datasets
• A publication model for data selection and storage
• Lexicons of controlled vocabulary terms supporting coordinated development of defined neuroscience attributes

• Hierarchic Attribute-Value implementation to facilitate inter-domain inquiries
• Multidimensional representations for interoperability

Reflecting the strong collaborative culture of the Human Brain Project, each of these has been adopted, or is being adapted, by other Human Brain Project resources, often with additional enhancements that aid our project as well.

• ***Commercial Partnership for Development of Neurodatabase Software***

[THIS THRUST IS FUNDED BY A TWO YEAR SBIR PHASE II GRANT FROM THE NIMH]

The database tools designed in the Laboratory of Neuroinformatics allow query and display of archived datasets. The value of database records would be further enhanced by aiding the delivery of database data directly to standard applications used for neurophysiological data analysis, and providing as well plugins to facilitate data upload from standard data acquisition routines. To provide this capability, we have formed an alliance with Bruxton Corporation, a respected small-business developer of data analysis software for neuroscience and biophysics. Collaboratively, we are developing software tools to provide these data entry and analysis links to neuroscience databases via popular Bruxton and third-party applications. The business model provides that Bruxton will provide the user tools free of charge, relying on the increased utilization of their and third-party programs to generate income.

II: PLANNED FUTURE EXPANSION OF NEUROINFORMATICS

• ***Algorithms for Information Transmission in Cortical and Parallel Neuronal Networks***

[A NEW, ADDITIONAL, 4-YEAR HUMAN BRAIN PROJECT PROPOSAL TO SUPPORT THIS WORK HAS RECEIVED A PRIORITY SCORE IN THE 8TH PERCENTILE, AND HAS BEEN RECOMMENDED BY THE HBP FOR FUNDING AT NIH COUNCILS IN OCTOBER 2003.]

A new collaborative thrust will develop, implement, and apply parallelized computational algorithms to explore the information content of spike trains and other neuronal signals, towards an understanding of the neural coding underlying visual and somatosensory processing. This project brings to bear local and external collaborators, with local resources (our databases and 26-processor Beowulf array), to explore informational aspects of neural coding and processing. Via both user-specified and project-developed algorithms, we will enable analyses to be performed either on-the-fly during dataset submission, or on archived data, permitting post-hoc examination as well as searches for specific patterns of brain activity. A major goal is development of the new field of *computational neuroinformatics*

Believing strongly that informational, computational, or theoretical biology should never be divorced from experimental work, this thrust also continues my laboratory's long-standing interest in neural networks, their neuronal and synaptic components, and their emergent properties. Using techniques I developed and introduced for simultaneous voltage-clamping of multiple interconnected neurons, we will analyze the information carrying and processing capabilities of parallel channels formed by paired *Aplysia* neurons. These form a testbed, bridging the gap between the single neurons characteristic of invertebrates and the massively parallel columns and modules found in mammalian brains. Related experiments may test aspects of the "fire-together, wire-together" hypothesis. This work descends as well from studies of interneuronal organization begun 35 years ago in the laboratory of Eric R. Kandel.

• ***Collaborative Development of BrainML Lexicons***

[FOLLOWING PRELIMINARY WORK UNDER CURRENT FUNDING, A COLLABORATIVE MULTI-INSTITUTION PROPOSAL TO SUPPORT THIS WORK IS PLANNED FOR A FUTURE SUBMISSION.]

Lexicons for the cortical database are being populated under our current HBP grant. Lexicons are built with hierarchic controlled vocabulary within the syntactic framework of our BrainMetaL HAV schema, which defines trees of terms associated with specified attributes. Development of additional lexicons for expanding domains within neuroscience and biophysics will properly be carried out collaboratively with neurobiologists and others representing disparate areas, as well as experts in lexical informatics. We therefore expect to develop collaborative links for this work with medical informatics groups and others.

• ***NIH-Mandated Data Sharing Initiative***

[PRELIMINARY DEVELOPMENT WORK UNDER CURRENT FUNDING MAY BE EXTENDED VIA NIH ADMINISTRATIVE SUPPLEMENT, OR AS A SEPARATE PROJECT FOR A FUTURE FUNDING CYCLE.]

Our laboratory's work supports the goals of data sharing, and develops methodologies that allow neuroscience data to be shared. As a community resource to present and explore technological and sociological issues affecting data sharing, we have established the web site datasharing.net. Its content includes ethical perspectives relating to the interests and responsibilities of producers and sharers of data, and users of shared data.

In March 2003, the NIH released a final policy requiring all high-direct-cost grant proposals submitted in October, 2003 or after to specify a plan for sharing data acquired by the project. This imminent schedule does not provide time for most areas of neuroscience to develop centralized databases such as those we have provided under our initial phases. As a consequence, there is an urgent need for rapidly-deployable peer-to-peer networks for data sharing. The Laboratory is responding to the challenge presented by this new mandate not only by making current schemas, methods, and lexicons public, but also by the development of a new methodology for peer-to-peer serving and searching heterogeneous neuroscience data—GENIE—the Generalized Extensible Neuroscience Internet Examiner. GENIE is layered on BrainML, but can use other XML-based or free-text schemas for domains of neuroscience that have not yet had BrainML lexicons developed. We are likely to coordinate our efforts with two other HBP groups that are devising similar schemes, towards interoperability or possibly a synthesis.

FOLLOW OUR PROGRESS

General progress, as well as technical or normative specifications for our developing neuroscience and biophysics markup languages, may be tracked at our web sites:
http://neurodatabase.org
http://brainml.org
http://datasharing.net

Daniel Gardner, Ph.D.
Head, Laboratory of Neuroinformatics
Department of Physiology & Biophysics
Weill Medical College of Cornell University
1300 York Avenue
New York, NY 10021
USA

Voice: + 1 212 746 6373
Fax: + 1 212 746 8355
Email: dan@med.cornell.edu