Seminars in
Pediatric
Neurology

The NeuroDevNet Neuroinformatics Core

Elodie Portales-Casamar, PhD,* Alan Evans, PhD,† Wyeth Wasserman, PhD,‡ and Paul Pavlidis, PhD*

The field of neuroinformatics has expanded dramatically during the past decade building on the development of new technologies in brain research as well as in computing. The activities are diverse, from data management and standardization that has become essential due to the large amount of data generated and the needs to share it, to the development of sophisticated software necessary for the analyses and visualization of such data. NeuroDevNet is a Canadian initiative, funded by the Networks of Centres of Excellence, devoted to the study of brain development with the goal to translate this knowledge into improved diagnosis, prevention and treatment of neurodevelopmental disorders. The NeuroDevNet Neuroinformatics Core is dedicated to helping researchers across the network with their data management, standardization and sharing, as well as to implement innovative solutions to facilitate their research. It is an essential component to NeuroDevNet, enabling active collaboration across the country and optimizing this unique endeavor.
Semin Pediatr Neurol 18:17-20

Neuroinformatics is the application of informatics tools to better understand the brain. This relatively new field has an impact on many facets of neuroscience, from the analysis of the brain at the molecular level to the modeling of complex behaviors.[1-3] Neuroinformatics activities have recently multiplied in part thanks to initiatives such as the Human Brain Project,[4] the International Neuroinformatics Coordinating Facility,[5] the Biomedical Informatics Research Network,[6] and the Neuroscience Information Network.[7] For the purposes of this article, neuroinformatics activities can be classified into the following major categories, which are discussed in some more detail later: (1) databasing and organizing publicly available data, (2) developing means of data standardization and interoperability, (3) methods and tools for imaging informatics, (4) computational modeling of the brain and its components, and (4) neuroscience-oriented tools for genomics analysis.

Creating databases of publicly available data is a key activity in bioinformatics and plays a prominent role in neuroinformatics. Examples include SenseLab,[8] NeuroExtract,[9] CCDB,[10] BAMS,[11] CoCoMac,[12] fMRIDC,[13] neurodatabase,[14] and GENSAT.[15] Each of these databases attempts to fill a niche in the needs of neuroscientists by aggregating information in 1 place. The creation of simple data collections does not ensure that the data can be easily used. A great deal of attention has been paid to making data usable not just by humans browsing the web but by computational approaches that can leverage large quantities of data. To make the best use of the data, it must be represented in a carefully designed and standardized way using structured data representations and controlled vocabularies. For complex data types particularly (eg, images), descriptive information ("meta-data") need to be stored along with the raw data to allow users to readily search through large datasets. In addition, because of the wide variety and complexity of data types, one of the major roles of neuroinformatics is to ensure interoperability between resources by defining data-exchange standards and data integration over distributed resources. This activity is a constant challenge considering the heterogeneous data sources and the variety of available scientific terminologies.[16] A few examples of existing neurosciences-specific standards that strive to formalize these vocabularies include NeuroNames,[17] NeuroML,[18] the BrainML,[19] the NIFSTD, and BIRNLex vocabularies.[20]

Brain imaging is one of the focal points of neuroscience-specific computational developments. This is because brain imaging produces large and complex datasets that require complex statistical methods to extract useful information.[4,21] Technological advances, such as faster networks, improved data compression, and cluster-, grid- and cloud-computing, allow for easier management and analysis of these large files. The Biomedical

From the *Centre for High-Throughput Biology, University of British Columbia, Vancouver, BC, Canada.
†Montreal Neurological Institute, McGill University, Montreal, QC, Canada.
‡Center for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada.
Address reprint requests to Elodie Portales-Casamar, PhD, Centre for High-Throughput Biology, University of British Columbia, 2125 East Mall, Vancouver, BC, V6T1Z4, Canada. E-mail: elodie@neurodevnet.ca

1071-9091/11/$-see front matter
doi:10.1016/j.spen.2011.02.001

Informatics Research Network exemplifies a large-scale initiative providing a grid-based framework to researchers for the management and analysis of neuroimaging data.[6] In regard to image data standardization, several computer-based atlases have been developed that allow for the mapping of brain structures to a common 3-dimensional coordinate system (http://www.loni.ucla.edu/Atlases for a listing). This indexing of the brain allows for the development of algorithms to normalize, visualize, and process brain-imaging data in an automated way. The neuroimaging informatics tools and resources clearinghouse is a repository of neuroimaging tools and resources.[22]

Another active domain in neuroinformatics is mathematical modeling. Many aspects of brain function can be modeled, including, but not limited to, a single neuron,[23] spiking neural networks,[24] and large-scale patterns of activity.[25] This activity requires a good understanding of the structure of the underlying system and is in some ways still in its infancy.

Finally, the revolution of genome sequencing and functional genomics has had a major impact in the neuroinformatics field because the discoveries were applied to the study of the brain. The Allen Brain Atlas provides a mapping of gene expression to the mouse, and more recently human, brain regions.[26] The GeneNetwork enables researchers to infer linkage between gene variants and complex traits.[27] Gemma is a framework for the analysis of gene-expression data collected for public microarrays, including many brain-related datasets.[28]

The NeuroDevNet Neuroinformatics Core

The brief background in neuroinformatics given earlier sets the stage for discussing the role of neuroinformatics in NeuroDevNet. A major purpose of the NeuroDevNet Networks of Centres of Excellence (NCE) is to facilitate collaborative research in brain development in Canada. For collaborative research to occur, information must be exchanged. Although traditional means of communication are necessary, the interdisciplinary nature of translational research, the complexity of data generated by cutting-edge genomics and imaging techniques, and the availability of high-performance data networks and computational infrastructure present outstanding opportunities to do more. Essential to NeuroDevNet is the integration of a Neuroinformatics Core for the implementation of innovative solutions for research collaboration within the network.

We see the Neuroinformatics Core constituting a "glue" that binds the many aspects of NeuroDevNet together. It speeds research by increasing communication and use (including reuse) of data, results, tools, and ideas. It plays largely a coordinating role rather than embarking on major infrastructure building. This allows us to benefit from existing expertise in the network and reuse existing infrastructure and resources, avoiding a long ramp-up time before it delivers value to NeuroDevNet.

Informatics Support for Data Management and Sharing Across Multiple Sites

Each NeuroDevNet project includes scientists charged with the generation, capture, and analysis of data. These individuals delve deeply into the data, create productive collaborative interactions, and ultimately provide the intellectual catalyst to extract meaning from pools of data. The Neuroinformatics Core serves the needs of these individuals, capitalizing on opportunities for interactions and exchanges among them. Where suitable, the innovative methods and software used within one project are made available to the other projects.

The task of coordinating the exchange of information across far-flung researchers is a significant challenge. In principle, 1 option might be to centralize the myriad data generated by NCE laboratories. This is not feasible because of time, cost, technical challenges, patient privacy, and data-ownership concerns, not to mention the fact that most groups already have established some informatics capacity to address local needs. A second option is to create a "federated data grid" in which each NCE laboratory implements a common software interface around its databases that can be accessed seamlessly by investigators (with appropriate permissions) at any site. This model has been adopted by caBIG (the cancer informatics program of the US National Cancer Institute), but development took many years and many millions of dollars. Neither of these models suits our purposes. Therefore, we have determined that the Neuroinformatics Core must adopt a simpler plan to succeed. We are implementing, where possible, data and tools that interoperate across NeuroDevNet laboratories but without embarking on vast software and database development projects. We refer to this as a "confederation coordination" model. Infrastructure is provided to permit the exchange of information across the network and supported by the Neuroinformatics Core to reduce the effort needed by the researchers to provide information and to use that provided by others.

NeuroDevNet researchers work on a wide variety of subjects and use many different methods that generate distinct types of data (from rat behavioral tests to human magnetic resonance imaging (MRI) through genetics analyses). Because no existing software can handle all of these different data types, the Neuroinformatics Core needs to identify adequate software for each data type separately and then develop data-exchange solutions between the different software. The use of software for data management enables the implementation of standards as described in the Introduction. In the specific case of the NeuroDevNet demonstration projects, standards for neurologic examination and diagnosis have already been developed by experts in each field. This ensures standardization of the clinical data within each project. The Neuroinformatics Core is working on homogenizing these standards across projects to ease the comparison of studies across neurodevelopmental disorders. Two software programs identified by the NeuroDevNet Neuroinformatics Core as suitable for the NeuroDevNet requirements for clinical and imaging data, respectively, are presented later.

Research Electronic Data-Capture Software for Clinical Research Data Management

All the NeuroDevNet demonstration projects have a clinical data-collection component and require appropriate software to facilitate this effort in a standardized way across multiple sites. For instance, one of the cerebral palsy (CP) demonstration project's major activity is to develop a national population-based registry to better understand CP etiology in the Canadian population and its risk factors. Current practices in the CP registry do not meet the demands of a modern multisite project designed to take the study of CP into the next stages of analysis discovery. Specifically, there is a demand for a more powerful database and software solution to collecting subject data as well as to facilitate tracking across sites.

Multiple software programs are available to manage clinical data (eg, Dacima, http://www.dacimasoftware.com; OpenClinica, http://www.openclinica.org; and REDCap, http://www.project-redcap.org). The Neuroinformatics Core is committed to use, whenever possible, open-source and/or community-based software solutions to minimize the cost for long-term maintenance. With this in mind, we elected the research electronic data-capture (REDCap) framework as the best solution for clinical data management (http://project-redcap.org/). REDCap is a case report forms platform developed by Vanderbilt University. The software is available at no cost for REDCap Consortium Partners, which NeuroDevNet has recently become. The Neuroinformatics Core has identified the opportunity of developing this activity through a partnership with the Maternal Infant Child and Youth Research Network and the Clinical Research Informatics Centre at the Women's and Children's Health Research Institute in Edmonton who are actively maintaining clinical databases for the research groups in their institution.

We are in the final stage of implementing the CP registry using REDCap. The Neuroinformatics Core is working with the CP project team to develop the forms, focusing on consistent data representation and software reusability. The experience gained in establishing the CP Registry can be applied to other clinical data-collection activities throughout the NeuroDevNet network.

LORIS Framework for Behavioral and Brain-Imaging Data Management

The Neuroinformatics Core is taking advantage of a powerful web-based brain-imaging database architecture (LORIS) and analysis pipeline (CIVET) that was developed at the Montreal Neurological Institute by Dr Alan Evans' group to underpin the $30 million multicenter National Institutes of Health MRI Study of Normal Brain Development. Within the National Institutes of Health MRI Study of Normal Brain Development project, repeated studies in each child allow for both longitudinal and cross-sectional analysis. Imaging and extensive (>10,000 fields) behavioral data are transferred via a web-based network into a database that allows for (1) secure, encrypted data transfer and automated quality control; (2) automated large-scale MRI segmentation; (3) correlation of neuroanatomic and behavioral variables as 3-dimensional statistical maps; and (4) remote interrogation and 3-dimensional viewing of database content. The LORIS database has since been used to support other multicenter network projects, studying both developmental and neurodegenerative disorders. For example, the LORIS database supports the Infant Brain Imaging Study, a multicenter US study of autism. LORIS is also being used to enable the $4 million CIHR-funded MAVAN project, a collaboration between 3 Canadian centers in which Dr Michael Meaney and colleagues are examining whether aspects of parental care, such as touch, have the same effects on human babies as shown in rat models. LORIS and CIVET are being deployed as part of the $2.4 million CANARIE (Canada's advanced research and innovation network)-funded Network-Enabled Platform project, Canadian Brain Imaging Research Network. CANARIE is an initiative from the government of Canada to promote collaboration in research, innovation, and higher education through the development of digital technologies. The Canadian Brain Imaging Research Network, now extended to an international initiative (Global Brain Imaging Research Network), will create a service-oriented architecture to link Canadian brain-imaging research centers to high-performance computing resources in Canada and internationally. Successful deployment of this infrastructure will provide NeuroDevNet with a natural platform for the joint collection and sharing of imaging/behavioral/genetic data across multiple sites. This will be particularly important for the coordinated management of a multisite data being collected according to a uniform acquisition protocol as well as to provide computing resources that are often a limiting factor when working with large data files, such as brain images. LORIS will be applied to all 3 NeuroDevNet demonstration projects.

We stress that LORIS and the REDCap are just examples of existing tools NeuroDevNet can support to facilitate the research goals of the demonstration projects. The Neuroinformatics Core will play a role in helping network members identify relevant tools and assisting in their use. Resources will be provided to subsidize the development, adaptation, installation, and/or purchase of such tools.

Tracking Efforts and Results Across and Beyond the Network

One of the major challenges in the administration of a network such as NeuroDevNet is communication. It is essential that all efforts and discoveries are seamlessly communicated across the network and eventually to the public. The Neuroinformatics Core strives to make the NeuroDevNet network a strong entity by building on each other's strength, avoiding duplication of efforts, and working as a team on protocol and standard establishments.

Helping With Communication and Collaboration

The NeuroDevNet web site is the main platform to achieve this goal, with a public section maintained by the NeuroDevNet communication officer and an internal section for member use only developed by the Neuroinformatics Core. The latter section is meant as a hub to facilitate the exchange of data, tools, ideas, and other resources across NeuroDevNet and to play additional roles in the organization of NeuroDevNet activities. The Neuroinformatics Core has adopted a wildly used team collaboration tool, "confluence" (Atlassian, Inc, Sydney, Australia; http://www.atlassian.com). Confluence is a wiki (editable web site) that has extensive support for activities, such as networking, collaboration, document sharing, project management, and so forth. Each demonstration project has its own wiki space where they can share project-specific information in a private manner; additional wiki spaces are used by the cores and the NeuroDevNet management team to distribute information to all NeuroDevNet members. In addition to the wiki, we are implementing a multiuser file repository, the Owl intranet engine (http://owl-.anytimecomm.com), to ease the sharing of large data files (up to 2 Gb). In this repository, files can be shared across the whole network or only a restricted group of people.

Meta-data Repository/Clearinghouse

Although the Neuroinformatics Core is not creating a centralized warehouse for all data collected by NeuroDevNet sites, we are creating a database to track meta-data (descriptive information about data), tools, reagents, and protocols. We call this the "clearinghouse" although it emphasizes meta-data and not raw research data. No data will be stored that would raise privacy or ethics questions that have not already been addressed (or which can be easily addressed) by the data providers. For example, each project will provide details about the types of subjects studied, which assays have been run, the formats in which the data are available, and where the data are kept. We are not asking projects to provide any actual data other than basic demographic information, so data ownership, privacy, and ethics issues are kept to a minimum. To facilitate this task, a secure web-based software system will be provided to solicit the essential information, accelerate the deposition process, and perform quality assessments of the meta-data. Updates of details of the studies will be done using automated tools we will develop.

Research in Neuroinformatics

The primary goals of the Neuroinformatics Core are as outlined previously, providing support and training "services" to NeuroDevNet. However, as described in the Introduction, neuroinformatics is a dynamic research discipline in which the leaders of the Core are actively engaged. This includes the development of methods for the management, analysis, and interpretation of data (including genomics, imaging, clinical, and behavioral data). This type of research falls outside the strict scope of the Core as a service provider and will not be discussed in detail here. However, the commitment of the Neuroinformatics Core scientists is motivated by the opportunity to engage in collaborative neuroinformatics research. Some examples include the selection of candidate genes, machine-learning analysis of behavioral data, analysis of regulatory networks in the developing brain, and interpretation of high-throughput sequencing data across NeuroDevNet projects.

References

1. French L, Pavlidis P: Informatics in neuroscience. Brief Bioinform 8:446-456, 2007
2. Martone ME, Gupta A, Ellisman MH: E-neuroscience: Challenges and triumphs in integrating distributed data from molecules to brains. Nat Neurosci 7:467-472, 2004
3. Van Horn JD, Ball CA: Domain-specific data sharing in neuroscience: What do we have to learn from each other? Neuroinformatics 6:117-121, 2008
4. Brinkley JF, Rosse C: Imaging and the Human Brain Project: A review. Methods Inf Med 41:245-260, 2002
5. De Schutter E: The International Neuroinformatics Coordinating Facility: Evaluating the first years. Neuroinformatics 7:161-163, 2009
6. Bug W, Astahkov V, Boline J, et al: Data federation in the Biomedical Informatics Research Network: Tools for semantic annotation and query of distributed multiscale brain data. AMIA Annu Symp Proc 6:1220, 2008
7. Gardner D, Akil H, Ascoli GA, et al: The neuroscience information framework: A data and knowledge environment for neuroscience. Neuroinformatics 6:149-160, 2008
8. Samwald M, Chen H, Ruttenberg A, et al: Semantic SenseLab: Implementing the vision of the Semantic Web in neuroscience. Artif Intell Med 48:21-28, 2010
9. Crasto CJ, Masiar P, Miller PL: Neuroextract: Facilitating neuroscience-oriented retrieval from broadly focused bioscience databases using text-based query mediation. J Am Med Inform Assoc 14:355-360, 2007
10. Martone ME, Tran J, Wong WW, et al: The cell centered database project: An update on building community resources for managing and sharing 3D imaging data. J Struct Biol 161:220-231, 2008
11. Bota M, Swanson LW: Collating and curating Neuroanatomical Nomenclatures: Principles and Use of the Brain Architecture Knowledge Management System (BAMS). Front. Neuroinformatics 4:3, 2010
12. Kotter R, Maier J, Margas W, et al: Databasing receptor distributions in the brain. Methods Mol Biol 401:267-284, 2007
13. Van Horn JD, Grethe JS, Kostelec P, et al: The Functional Magnetic Resonance Imaging Data Center (fMRIDC): The challenges and rewards of large-scale databasing of neuroimaging studies. Philos Trans R Soc Lond B Biol Sci 356:1323-1339, 2001
14. Gardner D. Neurodatabase.org: Networking the microelectrode. Nat Neurosci 7:486-487, 2004
15. Gong S, Zheng C, Doughty ML, et al: A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature 425:917-925, 2003
16. Cimino JJ, Zhu X: The practical impact of ontologies on biomedical informatics. Yearb Med Inform 124-135, 2006
17. Bowden DM, Dubach M, Park J: Creating neuroscience ontologies. Methods Mol Biol 401:67-87, 2007
18. Gleeson P, Crook S, Cannon RC, et al: NeuroML: A language for describing data driven models of neurons and networks with a high degree of biological detail. PLoS Comput Biol 6:e1000815, 2010
19. Gardner D, Goldberg DH, Grafstein B, et al: Terminology for neuroscience data discovery: Multi-tree syntax and investigator-derived semantics. Neuroinformatics 6:161-174, 2008
20. Bug WJ, Ascoli GA, Grethe JS, et al: The NIFSTD and BIRNLex vocabularies: Building comprehensive ontologies for neuroscience. Neuroinformatics 6:175-194, 2008
21. Aw T: Neuroimage databases: The good, the bad and the ugly. Nat Rev Neurosci 3:302-309, 2002
22. Luo XZ, Kennedy DN, Cohen Z: Neuroimaging informatics tools and resources clearinghouse (NITRC) resource announcement. Neuroinformatics 7:55-56, 2009
23. Ichikawa K: A modeling environment with three-dimensional morphology, A-cCell-3D, and Ca2+ dynamics in a spine. Neuroinformatics 3:49-64, 2005
24. Brette R, Rudolph M, Carnevale T, et al: Simulation of networks of spiking neurons: A review of tools and strategies. J Comput Neurosci 23:349-398, 2007
25. He Y, Evans A: Graph theoretical modeling of brain connectivity. Curr Opin Neurol 23:341-350, 2010
26. Jones AR, Overly CC, Sunkin SM: The Allen Brain Atlas: 5 Years and beyond. Nat Rev Neurosci 10:821-828, 2009
27. Rosen GD, Chesler EJ, Manly KF, et al: An informatics approach to systems neurogenetics. Methods Mol Biol 401:287-303, 2007
28. French L, Lane S, Law T, et al: Application and evaluation of automated semantic annotation of gene expression experiments. Bioinformatics 25:1543-1549, 2009