Understanding the brain through neuroinformatics

Jan G. Bjaalie

Centre for Molecular Biology and Neuroscience & Institute of Basic Medical Sciences, University of Oslo, Norway
Correspondence: jan.bjaalie@incf.se

***Jan Bjaalie** is Executive Director of the International Neuroinformatics Coordinating Facility at Karolinska Institutet, Stockholm, Sweden, and Professor of Neuroscience at the Institute of Basic Medical Sciences, University of Oslo, Norway. He received his PhD degree in 1990, and was appointed Associate Professor at the University of Oslo in 1992 and full Professor in 1997. His research group joined the Centre for Molecular Biology and Neuroscience, a Norwegian Centre of Excellence appointed by The Research Council of Norway, in 2002. His laboratory has discovered fundamental principles of sensory map transformations in large projection systems of the brain and has developed tools and data systems for regional and whole brain atlasing and histological mapping.*

The grand challenge in *neuroscience* is—it would seem—the effort of understanding the structure, function, and development of the nervous system in health and disease. Such understanding requires the integration of huge amounts of heterogeneous and complex data collected at multiple levels of investigation. The interdisciplinary field of *neuroinformatics* combines neuroscience with information science/technology and deals with the creation of the data systems that will be required to achieve such integration. The grand challenge in neuroinformatics is to achieve advanced, ultimately seamless, integration of all data needed to understand the nervous system.

Integration of neuroscience data may be compared to a multi-dimensional puzzle. One dimension of the puzzle is time. Thus, the adult brain and each developmental stage hold their own set of structural and functional data. Another dimension is the multiple levels of investigation. The levels span from genes and molecules through synapses, cells, networks, regions, and whole brain, to cognition and behavior. Fitting data together across time and within and across the many levels of investigation is a gigantic challenge. From a neuroinformatics perspective, the challenge may be described as related to 1) accumulation, storage, management, and sharing of data—the "databasing challenge", 2) development and sharing of tools for data analyses—the "tools challenge", and 3) creation and validation of computational models of brain structure and function built on the available data—the "modeling challenge".

The databasing challenge: Each of the multiple levels of investigation generates vast amounts of primary data of many different formats. Challenges for database developments include the establishment of standards for ontologies, metadata descriptions, and data formats, and the creation of mechanisms (from technical to sociological and legal) for sharing the vast amount of data among researchers (Amari et al, 2002; Koslow, 2002; Eckersley et al., 2003).

Ontology is more than terminology. The ontology of a database is the definition of the elements and the relationships between the elements included in the database. It is difficult to arrive at a standard ontology but important to strive towards clarification of definitions and concepts, allowing data to be more easily compared and interpreted. *Metadata*, "data about data", are made up by data describing the primary research data.

Data collected in an experiment are only meaningful if a number of experimental conditions and parameters are provided. Research articles contain metadata but not in a structured format. Standards for minimal sets of metadata for experimental data are needed and can realistically be developed.

Many types of *data formats* are used in neuroscience. Standardized formats allow easier use of shared tools and comparison of data produced by different communities and laboratories. Standard *practices for sharing of data* are also important for the development of large databases.

With an increasing number of large databases, data sharing in neuroscience may gradually become as common and useful as it is in genomics, where the existence of very large bodies of data is leading to increased knowledge as well as products and services linked to the improvement of human health.

The tools challenge: Many techniques for visualization and analysis of data have been developed by other fields. But the neuroscience community is in need of special-purpose, optimized tools and algorithms. Data only make sense in the context of tools. Navigation and manipulation of data requires a multitude of tools. Integration builds not only on the accumulation of data within and across the many levels of investigation, but also on the tools used to compare data, create higher order representations, and extract principles. Over time, some of these tools may even benefit researchers in branches of the information sciences as they deal with issues related to brain function such as machine learning and robotic task planning.

The modeling challenge: As in all of science, the understanding of the systems and phenomena under study involves the development of models that are descriptive as well as predictive and explanatory. In neuroscience, the systems and phenomena are among the most difficult to model due to the many levels of investigation required to understand function and the complexities present at each level (Grillner et al. 2005). The only way to validate models of the sophisticated functions carried out by the nervous system is through confrontation with the data sets of neuroscience, using tools developed via neuroinformatics.

All of the above outlined challenges are demanding. Few neuroscience laboratories have the combined expertise to deal with all aspects of neuroinformatics: databasing, tools development and sharing, and modeling. Most neuroscience laboratories have only a few experienced researchers and carry a responsibility to train younger researchers. Data and tools sharing in the context of training is not an easy task. Performing the experiments, collecting the data, carrying out the initial analyses, and completing reports and publications often take most of the available time and resources. Standardization at all levels of investigation, additional data acquisition for completeness, and sharing may not be on the agenda. With the establishment of larger consortia of multiple research groups, as well as the establishment of new data production oriented institutions and services, neuroscience research is gradually changing. Such larger enterprises are developing an increasing number of data systems, standards, and requirements for populating the systems. With a growing number of available data systems covering one or several levels of investigation, it is to be expected that the neuroscience field will gradually move towards the use of such systems. The field of neuroinformatics will deal with the further development of the data systems and the challenge of making the systems work together—becoming interoperable.

Recently, new opportunities for international coordination in the field of neuroinformatics have emerged. With an international science policy mandate, multi-disciplinary and multi-national panels have analyzed the field and provided recommendations for its further development (OECD, 1999, 2002). This effort has provided a basis for global coordinated actions (for review, see Bjaalie and Grillner, 2007) within all of the three legs of neuroinformatics (databasing, tools development and sharing, and modeling), as outlined above. Neuroinformatics is a large area of multidisciplinary research poised to play an important role in supporting neuroscience in the information age.

REFERENCES

Amari, S-I., Beltrame, F., Bjaalie, J. G., Dalkara, T., De Schutter, E., Egan, G. F., Goddard, N. H., Gonzalez, C., Grillner, S., Herz, A., Hoffmann, K-P., Jaaskelainen, I., Koslow, S. H., Lee, S-Y., Matthiessen, L., Miller, P. L., Silva, F. M. D., Novak, M., Ravindranath, V., Ritz, R., Ruotsalainen, U., Sebestra, V., Subramaniam, S., Tang, Y., Toga, A. W., Usui, S., Pelt, J. V., Verschure, P., Willshaw, D., Wrobel, A., OECD Neuroinformatics Working Group (2002). Neuroinformatics: the integration of shared databases and tools towards integrative neuroscience. *J. Integrative Neurosci*. 1, 117–128.

Bjaalie, J. G., Grillner, S. (2007). Global neuroinformatics: the International Neuroinformatics Coordinating Facility. *J. Neurosci.* 27, 3613–5.

Eckersly, P., Egan, G. F., Amari, S-I., Beltrame, F., Bennett, R., Bjaalie, J. G., Dalkara, T., De Schutter, E., Goddard, N. H., Gonzalez, C., Grillner, S., Herz, A., Hoffmann, K-P., Jaaskelainen, I. P., Koslow, S. H., Lee, S-Y., Matthiessen, L., Miller, P. L., Silva, F. M. D., Novak, M., Ravindranath, V., Ritz, R., Ruotsalainen, U., Sebestra, V., Subramaniam, S., Toga, A. W., Usui, S., Pelt, J. V., Verschure, P., Willshaw, D., Wrobel, A., Yiyuan, T., OECD Working Group on Neuroinformatics (2003). Neuroscience data and tools sharing. A legal and policy framework for Neuroinformatics. *Neuroinformatics*. 1, 149–166.

Grillner, S., Kozlov, A., Kotaleski, J. H. (2005). Integrative neuroscience: linking levels of analyses. *Curr. Opin. Neurobiol*. 15, 614–21.

Koslow, S. H. (2002). Opinion: Sharing primary data: a threat or asset to discovery? *Nat. Rev. Neurosci.* 3, 311–3.

OECD (1999). Final report of the OECD Megascience Forum Working Group on Biological Informatics. http://www.incf.org/documents/incf-core-documents/Report_OECD_MSF_1999.pdf

OECD (2002). Report on neuroinformatics from the Global Science Forum Neuroinformatics Working Group of the Organisation for Economic Co-operation and Development. http://www.incf.org/documents/incf-core-documents/Report_OECD_GSF_2002.pdf

Front. Neurosci. (2008) 2, 1: 19–21. doi: 10.3389/neuro.01.022.2008