Neuroinformatics: definitions and data sharing policies

Erik De Schutter
Laboratory of Theoretical Neurobiology, University of Antwerp, Belgium
erik@bbf.uia.ac.be http://www.bbf.uia.ac.be

Neuroinformatics is growing field which originated in the nineties when both computing and large scale data storage became cheap while at the same time the world wide web was born in 1991. These converging trends led to a dream that all data acquired in neuroscience, most of which was now in digital format, could be made available over the web and be accessible for analysis with new sophisticated tools.

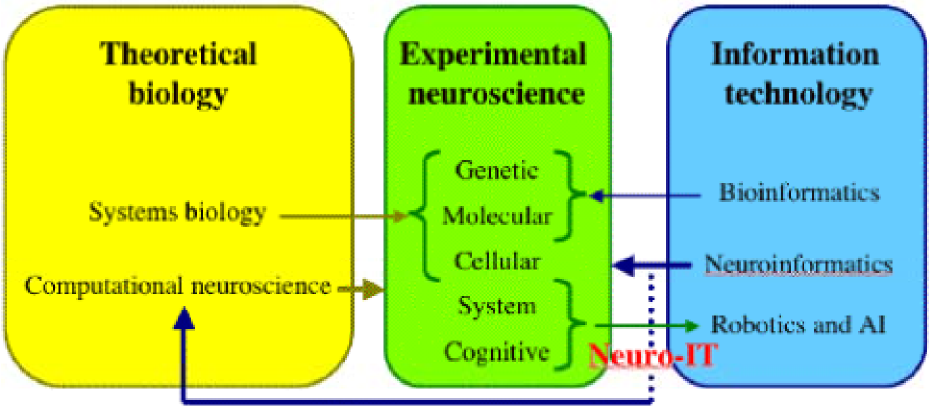
In 1993 the Human Brain Project was launched in the United States which provided a funding stream towards developing the informatics tools necessary to fulfill this vision and used the term neuroinformatics to capture this new kind of interaction between information technologies and neuroscience. Unfortunately the word neuroinformatics had been employed before in Europe to describe the study of information processing in the brain and several Neuroinformatics Institutes existed already.

These conflicting views forced the OECD working group on neuroinformatics, which was founded end of 1996 under the Megascience initiative, to spend more than a year on agreeing on a common definition which read "Neuroinformatics combines neuroscience and informatics research to develop and apply advanced tools and approaches essential for a major advancement in understanding the structure and function of the brain.". In the 2002 report generated by the continuation of the Neuroinformatics Working Group under the OECD Global Science Forum (Amari et al. 2003) this was further detailed as comprising three principal aims:

1. To optimise the accumulation, storage, and sharing of vast amounts of primary data and of large, structured neuroscience databases. As described above, the data are of an enormous diversity. The most immediate goal is to develop standards and mechanisms for sharing the vast amount of data among researchers.
2. To develop tools for manipulating and managing the data. Although many relevant techniques have already been developed in other fields, the neuroscience community must collectively design and develop special-purpose analytical tools and algorithms that are optimal for their needs. It is likely that, in the near future, large databases will play a similar role in neuroscience as they already do in genomics, where the existence of very large bodies of data, and of tools to navigate and manipulate these data, leads to breakthroughs in understanding and important commercial applications linked to human health. It is anticipated that some of these tools will, in turn, be of great benefit to researchers in various branches of the information sciences as they deal with problems (such as machine learning, robotic task planning, etc.) that are related to brain function in humans and other organisms.
3. To create computational models of brain structure and function that can be validated using the data. As in all of science, the understanding of the systems and phenomena under study involves the development of models that are not just descriptive but predictive and explanatory as well. In this case, the systems and phenomena are among the most difficult to model: from the molecular/cellular up to perception, learning, memory, reasoning, etc. The only way to validate models of these sophisticated phenomena is through confrontation with the data sets of neuroscience,

using tools developed via neuroinformatics.

While this definition has the advantage of fully covering neuroinformatics initiatives in Europe, Japan and the USA, it makes it more difficult to position neuroinformatics to related fields in biosciences like bioinformatics and systems biology. The tremendous recent growth in bioinformatics and the great interest in systems biology suggest that neuroscience would benefit from allying neuroinformatics to bioinformatics and computational neuroscience to systems biology.



Recently a new field was defined by the information technology branch of the European Commission: neuro-IT. Neuro-IT is defined as the interface between neuroscience and information technology where neuroscience teaches and inspires new, more efficient IT solutions, particular in the areas of robotics and artificial intelligence. While this possible application is alluded to at the end of the 2nd specific aim of neuroinformatics it is clearly not the focus of the field of neuroinformatics. In fact, simplifying, one could say that neuroinformatics and neuro-IT represent opposite interactions: one applies IT technology to neuroscience research, the other applies neuroscience knowledge to IT solutions.

When one considers neuroinformatics databases in more detail a variety of architectures and philosophies, each with its own merits and disadvantages, can be distinguished. One axis on which they may be considered is the degree of top-down control, or curation, involved in data entry. We have recently developed a grass-roots databasing approach (Cannon et al. 2002) which is one extreme of this scale in which there is no curation, minimal standardization and a wide degree of freedom in the form of records used to document data. Such a scheme has advantages in the ease of database creation and in the equitable assignment of perceived intellectual property by keeping the control of data in the hands of the experts who collected it. It does, however, require a more sophisticated infrastructure than conventional databases since the software must be capable of organizing diverse and differently documented data sets in an effective way. Recently a commercial software system has been developed which offers this functionality (http://www.axiope.com).

R.C. Cannon, F.W. Howell, N.H. Goddard and E. De Schutter: Non-curated distributed databases for experimental data and models in neuroscience. *Network: Computation in Neural Systems* 13: 415-428 (2002).

S.-I. Amari, F. Beltrame, J. G. Bjaalie, T. Dalkara, E. D. Schutter, G. F. Egan, N. H. Goddard, C. Gonzalez, S. Grillner, A. Herz, K.-P. Hoffmann, I. P. Jaaskelainen, S. H. Koslow, S.-Y. Lee, L. Matthiessen, P. L. Miller, F. M. D. Silva, M. Novak, V. Ravindranath, R. Ritz, U. Ruotsalainen, V. Sebestra, S. Subramaniam, Y. Tang, A. W. Toga, S. Usui, J. V. Pelt, P. Verschure, D. Willshaw and A. Wrobel: Neuroinformatics: The integration of shared databases and tools towards integrative neuroscience. *Journal of Integrative Neuroscience* 1: 117-128 (2003).