

Global Neuroscience: Distributing the Management of Brain Knowledge Worldwide

Giorgio A. Ascoli

Published online: 8 December 2012
© Springer Science+Business Media New York 2012

A frequent hot topic in neuroscience is the need for and challenge of thorough knowledge annotation.¹ Expressing knowledge in machine-readable format is a common goal of many if not all domains of modern science. The unsurpassed complexity of the nervous system makes this goal particularly arduous in neuroscience. The difficulty is amplified by the often-discussed multitude of relevant scales and data types,² from neurons³ to whole brain imaging,⁴ and from molecules to clinical applications.⁵

Continuously accelerating advances of computing technologies in the past two decades suggest that computational power may soon no longer be a limiting factor in the quest for a real-scale, biologically realistic model of an entire mammalian brain at the level of individual neurons. Parallel breakthroughs in high-throughput imaging and genetic approaches are fostering the emergence of several “big science” approaches for collecting comprehensive data sets of neural structure and function. Examples within the domain of “connectomics”

alone include the Human Connectome Project of the National Institutes of Health,^{6,7,8} the Fly Light Project of the Howard Hughes Medical Institute,⁹ and the Mouse Connectivity Atlas of the Allen Brain Institute,¹⁰ in addition to the grass root 1000 Connectomes Project,^{11,12} just to mention a few.

It is thus at least conceivable, if not likely, that in the not-so-distant future we will have access to gigantic databases of neural data as well as sufficiently powerful computers to model entire nervous systems. Although most neuroscientists would certainly welcome such a scenario, even at that point the problem of reverse engineering the brain would remain tremendously challenging. Why? Because building computational models based on experimental data requires those data to be annotated. In keeping with the connectomic theme, it is not sufficient to have a large connectivity matrix describing the blueprint of the entire circuit. Incorporating these data in a computational model requires explicit identification of the brain regions and neuron types for each of those synapses, so as to enable integration of appropriate biophysical details, sources of input, and targets of output.

Large data sets (which are inescapable given the complexity of the brain) imply huge annotation efforts. Extensive

¹ Akil, H., Martone, M. E., & Van Essen, D. C. (2011). Challenges and opportunities in mining neuroscience data. *Science*, 331(6018), 708–712.

² Bowden, D. M., Song, E., Kosheleva, J., & Dubach, M. F. (2012). NeuroNames: an ontology for the BrainInfo portal to neuroscience on the web. *Neuroinformatics*, 10(1), 97–114.

³ Hamilton, D. J., Shepherd, G. M., Martone, M. E., & Ascoli, G. A. (2012). An ontological approach to describing neurons and their relationships. *Front Neuroinform*, 6, 15.

⁴ Hsiao, M. Y., Chen, C. C., & Chen, J. H. (2011). BrainKnowledge: a human brain function mapping knowledge-base system. *Neuroinformatics*, 9(1), 21–38.

⁵ Lowe, C. R. (2011). The future: biomarkers, biosensors, neuroinformatics, and e-neuropsychiatry. *International Review of Neurobiology*, 101, 375–400.

⁶ neuroscienceblueprint.nih.gov/connectome

⁷ Toga, A. W., Clark, K. A., Thompson, P. M., Shattuck, D. W., & Van Horn, J. D. (2012). Mapping the human connectome. *Neurosurgery*, 71(1), 1–5.

⁸ Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E., Bucholz, R., et al. (2012). The human connectome project: a data acquisition perspective. *Neuroimage*, 62(4), 2222–2231.

⁹ janelia.org/team-project/fly-light

¹⁰ alleninstitute.org/science/public_resources/atlas/connectivity_atlas.html

¹¹ Biswal, B. B., Mennes, M., Zuo, X. N., Gohel, S., Kelly, C., Smith, S. M., et al. (2010). Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10), 4734–4739.

¹² Kennedy, D. N. (2010). Making connections in the connectome era. *Neuroinformatics*, 8(2), 61–62.

G. A. Ascoli (✉)
Krasnow Institute for Advanced Study, George Mason University,
Fairfax, VA, USA
e-mail: ascoli@gmu.edu

automation and optimal ergonomic design can minimize human involvement, but will never eliminate it altogether, because annotating data is ultimately rooted in human understanding. The same applies to extracting appropriate knowledge from scientific publications through literature mining, or even only tagging them for the presence of such knowledge.¹³ Thus, even after all necessary data are collected (and published) and when suitable computing platforms are available, creation of a working model of the brain will still require an enormous quantity of annotation person-years. In the end, the main factor determining *when* such a comprehensive brain model is completed will be the number of willing and active annotators.

Annotation typically involves specific skills, such as matching microscopy image with standard atlases or determining whether an article contains data relevant to a given mechanism. These skills can often be acquired and improved with continuous practice, but sometime only require surprisingly modest (e.g. undergraduate-level) scientific expertise. Annotation by entry-level trainees poses an issue of quality control, but as in many scientific designs, noisy signals can be dealt with by replication. For example, every piece of data could be blindly assigned to two independent annotators, and discrepant annotations would be checked by a supervisor. These observations suggest that neuroscience annotation may be amenable to “crowdsourcing” or broad distribution to semi-expert contributors.¹⁴ Early examples of this approach include involvement of students in the creations of wiki pages in psychology classes or labs¹⁵ and the Neuroscience Wikipedia Initiative of the Society for Neuroscience.¹⁶

In many cases, training for proper annotation of a specific dataset or literature search can be administered remotely via recorded tutorials based on a representative sample of the data or corpus. Annotation itself can also most times be performed at a distance. Thus, the issue becomes the online availability of potential annotators. Internet penetration figures (the proportion of population using the internet) from public online statistics and Census agencies¹⁷ indicate that, at the beginning of 2012, still less than a third of the nearly seven billion human beings on the planet used the internet. The main cause for the vast majority of the excluded population is lack of infrastructure, but this situation is very likely to change soon. Satellite telecommunication is predicted to

reach worldwide wi-fi coverage in less than 10 years,¹⁸ and the cost of portable devices for digital navigation is continuously dropping.

Tripling the number of potential annotators by global internet connectivity may not on the surface appear sufficient to solve the problem of neuroscience metadata. Social and economic considerations, however, are also at play. Recruiting, training, and retaining annotators in the developing World would be much more cost-effective than in the United States or in the European Community. In practice, distributing the task of neuroscience annotation to the entire human population could increase the scientific yield by one or two orders of magnitude. Consider a system in which laborious dataset annotation needs or difficult literature searches are posted online along with training material. Any willing contributor could undergo the online training. Once reaching a predetermined performance level, ‘certified’ annotators would gain access to the actual datasets. Subsequent data processing could then be compensated by the byte with sufficient wages to ensure comfortable standard of living for full-time annotators. Massive datasets will probably remain the purview of wealthier countries, while poorer countries would likely contribute most annotators.

Such a scenario would yield an accumulation of internationally shared data annotation; it would also result in a net flow of know-how (in the form of annotation training) and money (compensation) from the industrialized World to the developing countries. In the process, annotators would build specific and practical expertise, therefore gradually increasing their “market” value. Although some might consider such global knowledge distribution too utopian, an initial implementation of these ideas might be facilitated by the convergence of interests of major international neuroscience institutions. For example, the International Brain Research Organization (IBRO) has a core mission to provide for education and information dissemination relating to brain research, and to promote international collaboration and interchange of scientific information on brain research throughout the world. IBRO implements its mission by organizing schools and training programs across all continents, with a strategic focus on bringing neuroscience to the developing world. In countries without adequate infrastructures, setting up an annotation-centric neuroinformatics lab seems like a mild challenge compared to the heroic efforts required for enabling traditional wet-bench neuroscience. The initiative could be further aided by the involvement of the International Neuroinformatics Coordinating Facility (INCF), whose mission¹⁹ is indeed “to foster scientific interaction for discovery and innovation, and to facilitate the flow of information [...] to serve as a sustainable global

¹³ Ascoli, G. (2012). Twenty questions for neuroscience metadata. *Neuroinformatics*, 10, 115–117.

¹⁴ Hand, E. (2010). Citizen science: people power. *Nature*, 466(7307), 685–687.

¹⁵ psychologicalscience.org/index.php/publications/observer/2011/February-11/harnessing-the-power-of-wikipedia-for-scientific-psychology-a-call-to-action.html

¹⁶ sfn.org/index.aspx?pagename=neuroscienceQuarterly_09spring_wikipedia

¹⁷ Internet Usage and World Population Statistics. Internet World Stats (July 29, 2012). internetworldstats.com.

¹⁸ wi-fi.org

¹⁹ incf.org/about/what-we-do/our-mission

network for [...] internationally coordinated neuroinformatics activities and infrastructures [...; and t]o facilitate training of highly skilled neuroinformatics researchers worldwide.”

In the absence of such an internationally distributed effort, ongoing large scale neuroscience programs are each tackling their necessary annotation internally, incurring considerable additional cost besides data acquisition. In order to prepare the grounds for future high-throughput data production initiatives, the envisioned an-

notation system could be started and tested by recruiting and training annotators to add relevant metadata to PubMed abstracts. A practically suitable pilot project might focus on enhanced descriptions of electrophysiological studies using recommended Minimum Information Standards.²⁰ In turn, such an effort could feed back into existing neuroinformatics projects such as the Neuroscience Information Framework and the Neuron Registry,²¹ becoming immediately useful to modelers and experimental neuroscientists alike.

²⁰ Gibson, Frank, Overton, Paul, Smulders, Tom, Schultz, Simon, Eglén, Stephen, Ingram, Colin, Panzeri, Stefano, Bream, Phil, Sernagor, Evelyne, Cunningham, Mark, Adams, Christopher, Echtermeyer, Christoph, Simonotto, Jennifer, Kaiser, Marcus, Swan, Daniel, Fletcher, Marty, and Lord, Phillip. Minimum Information about a Neuroscience Investigation (MINI) Electrophysiology. Available from Nature Precedings <<http://hdl.handle.net/10101/npre.2008.1720.1>> (2008)

²¹ Hamilton, D. J., Shepherd, G. M., Martone, M. E., & Ascoli, G. A. (2012). An ontological approach to describing neurons and their relationships. *Front Neuroinform*, 6, 15.