

Value Added by Data Sharing: Long-Term Potentiation of Neuroscience Research

A Commentary on the 2007 SfN Satellite Symposium on Data Sharing

Yuan Liu · Giorgio A. Ascoli

© Humana Press Inc. 2007

Neuroscience research is resulting in an enormous accumulation of complex data. The explosive growth of bioinformatics demonstrated that sharing and reanalysis of digital information can lead to exciting opportunities for scientific integration and discovery. In striking contrast, broad-spanning issues (data heterogeneity, privacy regulations, etc.) have so far limited the reanalysis of brain and neural data. Nevertheless, data sharing is possible in neuroscience as well, and the ever-increasing power of affordable hardware and software has contributed to recent spiking interest in this topic. The Society for Neuroscience (SfN) has recently established a Neuroinformatics Committee and launched the Neuroscience Database Gateway, which is already widely used by the neuroscience community (Gardner and Shepherd 2004). Funding agencies, such as the National Institutes of Health (NIH), have data sharing policies, several journals require data deposition at the time of publication, and a number of books, journals, and reviews have focused on this topic in the past few years (e.g., Eckersley and the OECD Working Group on Neuroinformatics 2003).

For the 2007 SfN Annual Meeting, we have organized a Satellite Symposium entitled “The Rhyme and the Reason of Data Sharing” to present a series of recent success stories on data sharing and reuse in neuroscience. With examples

covering a vast array of techniques, topics, and scales, the symposium was designed to appeal to both neuroscientists who could share their data and to those who could reuse and reanalyze shared data. The scientific session consists of six short presentations ranging from molecular approaches to whole-brain imaging (described in more detailed below, in the order of presentation) followed by a panel discussion. The symposium is sponsored by four NIH Institutes (NINDS, NIMH, NIDA, and NCRR) and we have secured participation commitment of several prominent neuroscience leaders, including David van Essen, SfN President, and Tom Insel, NIMH Director (delivering the opening remarks), Nora Volkow, NIDA Director (moderating the panel discussion), and Story Landis, NINDS Director (providing the closing remarks).

Starting with “Multiple Reuses of *in vivo* Multi-Unit Recording Data”, the focus of Ken Harris (Rutgers University, Newark, NJ) is on electrophysiological time series (Harris 2005). Modern neurophysiological technology allows for simultaneous recording of hundreds of neurons, offering an unprecedented opportunity to study the function of neural circuits. The richness of this data (gigabytes per day in each lab) has fundamentally changed the work of physiologists. While a few decades ago the bulk of time was spent at the bench, now a single experiment is typically followed by months of computational analysis. Moreover, one data set can be suitable to address multiple scientific questions, which are often quite different from those that prompted the original experiment. In a recent example, Steuber et al. (2007) reanalyzed *in vivo* recordings in awake animals of other investigators (Goossens et al. 2004) to provide supporting evidence that the mechanisms they described using modeling and slice experiments also had behavioral consequences. More generally, strategies encouraging theorists to mine existing data

Y. Liu (✉)
NINDS, NIH,
Bethesda, MD, USA
e-mail: liuyuan@ninds.nih.gov

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

will benefit the progress of science. Harris took a leadership role in several successful case studies of electrophysiological data reanalysis, which lead to important insights on the social issues that need to be addressed to make large-scale data reuse a reality.

Shifting from functions to genes, “Reanalyses of Brain Gene Expression Data” by Paul Pavlidis (University of British Columbia, Vancouver, Canada) is also the topic of his Mini-Review in this issue of Neuroinformatics. As gene expression data rapidly accumulates in public databases, reanalysis has become common, and the value of sharing microarray data is increasingly appreciated. Pavlidis’ presentation aims at covering the development of databases and software systems for reanalyzing microarray data, to end with illustrative examples of analyses in the mammalian nervous system (e.g., Lee et al. 2004; Paterlini et al. 2005). The initial focus on building coexpression networks emphasized the ability of microarrays to generate hypotheses about gene function. Analyzing brain gene coexpression in a meta-analytical framework greatly increases predictive power. These techniques and tools (<http://www.bioinformatics.ubc.ca/Gemma>) can be applied to predicting function for poorly annotated genes in the nervous system using public data.

The topic tackled by Alumit Ishai (University of Zurich, Switzerland), “Mapping the Human Brain: New Insights from fMRI Data Sharing”, is further extended in her Mini-Review (with Jack van Horn) in this issue. The fMRI Data Center, established at Dartmouth in 2000, enables the open sharing of neuroimaging data from peer-reviewed papers to foster progress in cognitive science. The seminal study on the “Representation of objects in the human occipital and temporal cortex” (Ishai et al. 2000) marked the first deposition in the new database. Despite initial concerns about fMRI data sharing, this data set was frequently downloaded. Having been the author of the original work, describing the detailed brain activation patterns elicited by faces and objects in the human visual system, as well as of more recent developments (Fairhall and Ishai 2007) provides a unique perspective to overview several secondary analyses by independent investigators. A philosopher tested Husserl’s temporal components of consciousness. Other brain imagers deployed new analytic tools, from Dynamic Causal Modeling, estimating the neural interactions between cortical regions, to a novel method for constructing reproducibility maps. These re-analyses revealed new findings not reported in the original study, provided new perspectives on visual perception, generated new predictions, and resulted in new collaborations and publications in high profile journals (<http://www.ini.unizh.ch/~alumit/DataSharing.html>).

The talk by Richard Gilbertson (St. Jude Children’s Research Hospital, Memphis, TN), “Hunting for Cells of Origin of Brain Tumors with GENSAT—Gene Expression

Nervous System Atlas”, veers back to the molecular arena, with direct clinical implications. GENSAT is designed to document the expression patterns of all genes in the nervous system and to generate transgenic mice expressing reporter constructs that recapitulate the authentic patterns of selected genes. A major component of GENSAT, the Brain Gene Expression Map (BGEM) is a growing collection of in situ hybridization images of the developing and adult mouse (<http://www.stjudebgem.org>). A major interest of Gilbertson’s is in the use of BGEM to map the cells of origin of ependymoma, a brain tumor afflicting both children and adults. The analysis of 77 genes that distinguished cerebral, cerebellar and spinal ependymomas revealed dramatic spatial and temporal differentiation during development, and implied radial glia (RG) as a key expression site (Taylor et al. 2005). Co-immunofluorescence of fresh human ependymomas confirmed that these tumors contain rare fractions of cancer stem cells with RG-like morphology and immunophenotype. GENSAT transgenic reporter mice are now being used to study the susceptibility of RG to transformation by ependymoma-specific gene mutations and to similarly identify the neuronal cell lineages giving rise to other brain tumor types.

Offering insightful perspectives on “Collaborative Data Sharing between Neurophysiologists and Theorists”, Judith Hirsch (USC, Los Angeles, CA), moves the topic back to the neurophysiological arena (Hirsch and Martinez 2006). Hirsch has relevant experiences of her own experimental laboratory to share as well as of other physiologists who collaborate with theoretical neurobiologists. Ongoing discussions and computational re-analysis of intracellular data led to new theory-based projects. Slight modifications in experimental design permitted virtually simultaneous testing of both the lab’s staple projects and other independent ideas, emphasizing that the value of shared data can be augmented by flexible planning. A dialog between theorists and experimentalists requiring commitment from both sides is essential to educate each other about caveats and strengths of the respective approaches. While neuroanatomical reconstructions proved easier to share, physiological records often cannot be interpreted at face value, but necessitate much background and detailed knowledge of the individual data sets. Those interested in interdisciplinary collaborations might thus benefit from databases of scientists with complementary interests (a “match.com” for scientific laboratories) in addition to venues that bring together experimentalists and theorists.

The short series is closed by Giorgio Ascoli (George Mason University, Fairfax, VA) with “Successes and Rewards in Sharing Digital Reconstructions of Neuronal Morphology”, a topic also expanded in the related Mini-Review in this Neuroinformatics issue. The 3D reconstruction of neuronal arborization is a labor-intensive endeavor

carried out in many anatomical, electrophysiological, developmental, and pharmacological studies. Neuromorpho.Org is a SfN-endorsed and NIH-contracted repository of single-neuron reconstructions, publicly distributing thousands of cells from dozens of laboratories, brain regions, and morphological classes (<http://neuromorpho.org>) (Ascoli 2006). Ascoli's presentation of success stories highlights recent applications of these data to a variety of novel statistical analyses and computational models. The resulting fundamental discoveries include the quantitative characterization of potential synaptic connectivity in cortical circuits, the demonstration of morphological homeostasis in dendrites (Samsonovich and Ascoli 2006), the structural modulation of pyramidal cell excitability, and the functional processing ability of oblique branches in apical trees.

Finally, the panel discussion "Data sharing beyond just published" was envisioned to cap, complement, integrate, and critically extend the talks with crucial questions, such as "Why (and why not) share?", "What to share?", "When, and with whom to share?", "How to share?", and "What are the technical and sociological obstacles in data sharing, and how do we overcome them?". The Satellite Symposium is expected to be well matched with the Neuroinformatics Special Interest Social at the same 2007 SfN, bringing together four speakers for short talks and social interactions aimed at advancing neuroinformatics. The organizer of that event, Dan Gardner (Cornell University, New York, NY), will also present the Neuroscience Information Framework, a contract supported by the NIH Neuroscience Blueprint (<http://www.neuroscienceblueprint.nih.gov>), to be opened in beta form to be used and tested by SfN members. The neuroscience community is progressively becoming more open to data sharing, and it is hoped that initiatives such as these will help to accelerate and strengthen this process.

Acknowledgements The authors are indebted to the NIH Institutes sponsoring the Satellite Symposium and to the speakers, whose original abstracts have been worked into this commentary. The views

expressed in this commentary are those of the authors and do not represent those of the National Institutes of Health or the U.S. Government. GAA is supported by NIH grants NS39600, AG025633, and DA-HHSN271200577531C.

References

- Ascoli, G. A. (2006). Mobilizing the base of neuroscience data: The case of neuronal morphologies. *Nature Reviews Neuroscience*, *7*, 318–324.
- Eckersley, P., & the OECD Working Group on Neuroinformatics. (2003). Neuroscience data and tool sharing—A legal and policy framework for neuroinformatics. *Neuroinformatics*, *1*, 149–165.
- Fairhall, S. L., & Ishai, A. (2007). Effective connectivity within the distributed cortical network for face perception. *Cerebral Cortex*, in press.
- Gardner, D., & Shepherd, G. M. (2004). A gateway to the future of neuroinformatics. *Neuroinformatics*, *2*, 271–274.
- Goossens, H. H., Hoebeek, F. E., Van Alphen, A. M., Van Der Steen, J., Stahl, J. S., De Zeeuw, C. I., et al. (2004). Simple spike and complex spike activity of floccular Purkinje cells during the optokinetic reflex in mice lacking cerebellar long-term depression. *European Journal of Neuroscience*, *19*, 687–697.
- Harris, K. D. (2005). Neural signatures of cell assembly organization. *Nature Reviews Neuroscience*, *6*, 399–407.
- Hirsch, J. A., & Martinez, L. M. (2006). Circuits that build visual cortical receptive fields. *Trends in Neurosciences*, *29*, 30–39.
- Ishai, A., Ungerleider, L. G., Martin, A., & Haxby, J. V. (2000). Representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience*, *12*(Suppl 2), 35–51.
- Lee, H. K., Hsu, A. K., Sajdak, J., Qin, J., & Pavlidis, P. (2004). Coexpression analysis of human genes across many microarray data sets. *Genome Research*, *14*(6), 1085–1094.
- Paterlini, M., Zakharenko, S. S., Lai, W. S., Qin, J., Zhang, H., Mukai, J., et al. (2005). Transcriptional and behavioral interaction between 22q11.2 orthologs modulates schizophrenia-related phenotypes in mice. *Nature Neuroscience*, *8*, 1586–1594.
- Samsonovich, A. V., & Ascoli, G. A. (2006). Morphological homeostasis in cortical dendrites. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 1569–1574.
- Steuber, V., Mittmann, W., Hoebeek, F. E., Silver, R. A., De Zeeuw, C. I., Hausser, M., et al. (2007). Cerebellar LTD and pattern recognition by Purkinje cells. *Neuron*, *54*, 121–136.
- Taylor, M. D., Poppleton, H., Fuller, C., Su, X., Liu, Y., Jensen, P., et al. (2005). Radial glia cells are candidate stem cells of ependymoma. *Cancer Cells*, *8*, 323–335.