

The Missing Piece of the Puzzle: Neuroinformatics at the Bench

Erik De Schutter¹

Published online: 9 April 2015
© Springer Science+Business Media New York 2015

A recurring theme in our editorials has been the limited success of data sharing in practice: too few neuroscientists share their data or models.¹ A growing number of stakeholders have become concerned about promoting data sharing, as evidenced by a recent survey by a major publisher.² To improve data sharing, efforts have been made to encourage more scientists to contribute their data by giving them rewards, like the citable data paper,³ or by contacting them personally based on publications listing suitable data⁴ and publicly listing both available and unavailable data, including non-responders.⁵ Databases containing shared neuroscience data have been

made easier to use and find.⁶ But surprisingly little has been done to help motivated neuroscientists organize their data in such a way that it is ready to be shared,⁷ despite the huge opportunities for improvement.

It is a sad fact that - although most scientific data is acquired digitally - very few research laboratories in academic environments have adopted modern data management practices. Though it may seem an improvement to the scientist involved, there is really not much difference between having data in a bunch of poorly named files on a student's hard drive compared to having boxes with tapes or microscopy slides on a shelf. In both cases the data is not annotated, not searchable and poorly linked to the description of the experiment, which, even now, is usually still handwritten in a lab notebook. In some aspects the current situation is even worse because new problems like poor back-up and security policies and obsolescence of data formats have arisen.

The solutions to poor data management seem fairly simple and can be summarized as:

- annotate data at the time it is generated, with as much automation as possible; this also applies to processed data
- digitize everything, including especially the lab notebook
- store all data in a centralized database with proper back-up, archiving, access control, security, etc.

¹ David N Kennedy, "Where's the Beef? Missing Data in the Information Age," *Neuroinformatics* 4(4), 271–73, doi:10.1385/NI:4:4:271; Giorgio A Ascoli, "The Ups and Downs of Neuroscience Shares," *Neuroinformatics* 4(3), 213–16, doi:10.1385/NI:4:3:213; De Schutter, E. (2014). The Dangers of Plug-and-Play Simulation Using Shared Models. *Neuroinformatics* 12(2), 227–8. doi:10.1007/s12021-014-9224-7;

² <http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont>

³ David N Kennedy, Giorgio A Ascoli, and Erik De Schutter, "Next Steps in Data Publishing," *Neuroinformatics* 9(4), 317–20, doi:10.1007/s12021-011-9131-0; Krzysztof J Gorgolewski, Daniel S Margulies, and Michael P Milham, "Making Data Sharing Count: a Publication-Based Solution.," *Frontiers in Neuroscience* 7 (2013): 9, doi:10.3389/fnins.2013.00009.

⁴ Sumit Nanda et al., "Doubling Up on the Fly: NeuroMorpho.org Meets Big Data.," *Neuroinformatics* 13(1), 127–29, doi:10.1007/s12021-014-9257-y.

⁵ <http://NeuroMorpho.org>

⁶ <http://www.neuinfo.org>; Daniel Gardner et al., "The Neuroscience Information Framework: a Data and Knowledge Environment for Neuroscience," *Neuroinformatics* 6(3), 149–60, doi:10.1007/s12021-008-9024-z.

⁷ Kennedy, D. N. (2012). The Benefits of Preparing Data for Sharing Even When You Don't. *Neuroinformatics*, 10(3), 223–224. doi:10.1007/s12021-012-9154-1; Ruchi Parekh, Rubén Armañanzas, and Giorgio A Ascoli, "The Importance of Metadata to Assess Information Content in Digital Reconstructions of Neuronal Morphology," *Cell and Tissue Research*, 2015, doi:10.1007/s00441-014-2103-6.

✉ Erik De Schutter
erik@oist.jp

¹ Computational Neuroscience Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan

Note that the first two goals strongly affect the workflow during actual experiments at the bench, while the last goal is really a central facility that can be deployed at different levels of granularity (the lab, department, institute, ...). Few scientists worry about data annotation and, if it happens, it is either a feature of the capturing software - the most logical solution - or annotation is added at a later date when requested⁷. Which data are not yet digitized is very domain specific, but the boom of cheap sensors driven by the smartphone revolution makes it likely that everything will be recorded digitally soon. Setting up a quality database that can handle a large variety of data sets is a daunting prospect for the average scientist and should not be built from scratch. Instead the future should bring easy to install and use products comparable to, for example, iTunes to manage music and video collections. Note that putting data into the cloud is usually not a complete database solution. Though the cloud assumes backup and access control, it often is just a file repository (e.g., dropbox), lacking the essential indexing and search facilities that a real database offers.

The astute reader may have noticed that I did not mention lab notebooks in the preceding paragraph. This is because, although it has received relatively little attention in the neuroinformatics community, much recent progress has been made in electronic lab notebooks (ELNs). Several viable commercial solutions are now available that are suited for labs in academic environments⁸ and many labs are trying out generic products like Evernote.^{9,10} Many of the ELNs are built on the software infrastructure that gave us social media and therefore emphasize the collaborative and interactive aspects of research in a lab. For example, LabGuru¹¹ is a web-based environment that allows lab members with different roles to interact on scientific projects and, because one can access its functionality from mobile devices, it is easy to use at the bench. But while it has many interesting features, including ways to specify where samples and reagents are stored physically in a lab, it does not provide real data management. Until recently, database management was available only for expensive ELNs aimed at industry, like Accelrys ELN,¹² but it is

now being offered by RSpace ELN¹³ which, interestingly, builds on a project that was originally neuroinformatics.¹⁴

ELNs coupled to databases for local research projects may be the perfect tool to introduce naive scientists to modern data management, because they extend a familiar concept that is central to their daily research work. Nevertheless, despite the rapid progress of ELN development, their impact is still limited.^{10, 15} This reflects the rather conservative attitude of scientists towards anything that impacts lab culture, a poor training in basic computer science and the lack of perception of a problem. It is noticeable that in the survey mentioned earlier², only 20 % of participants responded that they lacked the resources needed to share their data while the overall lack of data management procedures suggests that, in practice, this is a much more widespread issue.

Returning to neuroinformatics, it is quite obvious that current ELN development is mainly geared towards molecular and pharmaceutical sciences. There are no neuroscience examples on the slick websites of the companies developing these products. Conversely, there are very few neuroinformatics projects that develop solutions aimed to be deployed by a research lab,¹⁶ the focus remains on top-down approaches that promote centralized databases and the development of tools needed for such databases, like ontologies and data standards. The ELN community, by contrast, seems to offer a more practical bottom-up solution that also makes the data actually available for effective sharing. For example, records can be ‘tagged’¹¹. This is really a form of data annotation but in an intuitive, non-standardized way. By providing good templates for many experimental protocols these tags become, however, standardized by default. It seems time for the neuroinformatics field to investigate whether the current ELN systems can fully serve our community. If so, we should strongly promote the use of ELNs in neuroscience labs. If not, we should encourage projects to extend ELNs to neuroscience because their potential to enhance data sharing is obvious.¹⁷

⁸ Jim Giles, “Going Paperless: the Digital Lab.,” *Nature*, January 26, 2012, doi:10.1038/481430a.

⁹ <http://evernote.com>. The pro and cons of using Evernote are discussed in Walsh, E., & Cho, I. (2013). Using Evernote as an Electronic Lab Notebook in a Translational Science Laboratory. *Journal of Laboratory Automation*, 18(3), 229–234. doi:10.1177/2211068212471834; <http://accelrys.com/materialsinnovation/?p=2271>;¹⁰

¹⁰ <http://depth-first.com/articles/2013/03/08/evernote-as-an-electronic-lab-notebook-eln/>

¹¹ <http://labguru.com>

¹² <http://accelrys.com>

¹³ <http://lab-ally.com/products/r-space-eln/>

¹⁴ Robert C Cannon et al., “Non-Curated Distributed Databases for Experimental Data and Models in Neuroscience,” *Network (Bristol, England)* 13(3), 415–28.

¹⁵ Melissa A Haendel, Nicole A Vasilevsky, and Jacqueline A Wirz, “Dealing with Data: a Case Study on Information and Data Management Literacy,” *PLoS Biology* 10(5), e1001339, doi:10.1371/journal.pbio.1001339.t001.

¹⁶ An exception is the Helmholtz project <https://www.dbunic.cnrs-gif.fr/documentation/helmholtz/overview.html>

¹⁷ I thank Giorgio Ascoli and Michele Giugliano for comments on earlier versions of this editorial.