Training for Data Science: Imagine There's no Countries

Giorgio A. Ascoli¹

Published online: 18 October 2017
© Springer Science+Business Media, LLC 2017

The brain-mind relationship constitutes arguably the greatest standing mystery for humankind, and perhaps one of the oldest too. Yet only recently the international community seriously began to confront the magnitude of the challenge. A number of large-scale programs around the world broke ground in the past few years,¹ aiming for breakthrough advancements in neuroscience. The overall focus of these efforts, including the BRAIN initiative in the US and the Human Brain Project in the EU, is on neurotechnology, neuroinformatics, and collaborative research.²

Understanding the brain will undoubtedly require not only big data,³ but also big crowds of skilled individuals able and willing to annotate, analyze, and interpret those data. I will optimistically assume that the data are coming – both because it seems reasonable given ongoing scientific progress, and because we may otherwise have to reconsider our hope to solve the neuroscience puzzle. After the data are collected and available in big brain-data repositories, how can we create useful knowledge from them? The

recipe is simple: we want the smartest and most qualified people at work.⁴

Based on the above premises, the scientific community and human society at large should have a vested interest in ensuring high quality training of the persons with the greatest potential to contribute to revolutionary discoveries. In what appears like a logical policy, most countries provide grants and fellowships to support the best and brightest of their scientists and trainees. But is this nation-by-nation approach really the most effective way to proceed? I maintain it might not be.

Suppose the US picks the most capable fraction of their budding researchers and gives them resources and incentives to study. The EU picks the top of theirs and China, India, Russia, Argentina, and Australia do the same. A simple spreadsheet calculation demonstrates that if these countries pooled their resources and *collectively* supported the top percentile of all scientists and trainees worldwide, on average more talent would be recruited for the same amount of investment.

While this argument holds in principle for all economic investment in research and development, it is particularly cogent in a scenario in which massive datasets sit in electronically accessible virtual silos. When only bits need to move, it matters not where the solution for their optimal sorting comes from.⁵ Thus, presuming key advances in neuroscience require a substantial component of information technology, computational simulations, statistical modeling, and machine learning, we should cast the net to recruit and secure the most talented researchers as wide as possible.

¹ Huang, Z.J., Luo, L. (2015). It takes the world to understand the brain. *Science*, 350(6256), 42–4.

² Olds, J.L. (2016). The rise of team neuroscience. *Nature Reviews Neuroscience*, 17, 601–2.

³ Landhuis, E. (2017). Neuroscience: Big brain, big data. *Nature*, 541(7638), 559–561.

✉ Giorgio A. Ascoli
ascoli@gmu.edu

¹ Bioengineering Department; Neuroscience Program; Center for Neural Informatics; Structures, & Plasticity; and Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA

⁴ Marder, E. (2015). Understanding brains: details, intuition, and big data. *PLoS Biol*, 13(5), e1002147.

⁵ Roskams, J., Popović, Z. (2016). Power to the people: Addressing Big Data challenges in neuroscience by creating a new cadre of citizen neuroscientists. *Neuron*, 92(3), 658–664.

If in 20 years I am diagnosed with Parkinson's disease, will it matter to me who invented the medicine that cures me or whether the discovery came from New York, Beijing, Delhi, or Rome? And when a new smartphone app warns you that an aneurism is forming in your anterior cerebral artery allowing you to get preventive remediation before it ruptures into a devastating stroke – will you care whether the app was developed in a Google think tank, in an Israeli start-up, or in an unglorified garage anywhere on planet Earth?

The globalization of the selection and support of the next generation of neuroscience Big Data miners does not entail a utopian symmetry of financial commitment. In other words, two countries may pool resources and maximize their return on investment even with uneven contributions of money. Likewise, the number of trained scientist of each nationality under such an arrangement is not necessarily proportional to resources contributed by their own fellow tax-payers. To clarify this point, consider an extreme case of a single country willing to embark in such a program. In the current system, that country finds and funds the *N* best scientists and trainees among its citizens. By allowing all trainees worldwide to compete for the same fellowships, regardless of citizenship, the *N* best will encompass, by the very nature of ranking, more talented individuals than the original pool.

Consider moreover that the majority of scientific investment comes from high-income countries where the cost of living (and thus an individual fellowship's price tag) is much greater than in the rest of the World. In the above scenario the putative single-bidding country would likely be able to support a larger number of trainees, further boosting the scientific throughput for the same invested capital. Of course in this model the most costly *data generation* would still remain the purview of the wealthiest countries. But at least with respect to data analysis and interpretation, or in other word *knowledge generation*, the proposed system would in principle level the playing field between high- and low-income societies.⁶ In practice, this would enable more effective

engagement of underutilized intellectual assets from resource-limited countries.

Currently both of the most prominent mechanisms in America for supporting graduate students in STEM and biomedical research (the National Science Foundation Graduate Research Fellowship Program⁷ and the National Institutes of Health National Research Service Awards⁸) are rigorously limited to US citizens or permanent residents. How absurdly detached from reality does this restriction appear every morning when I walk into my lab full of thriving doctoral trainees holding passports from India, Pakistan, Italy, Iran, and Spain (for full disclosure, I have also been the privileged PhD adviser of seven US graduates). Foreigners similarly constitute the overwhelming majority of the masters-level software programmers that keep NeuroMorpho.Org⁹ and Hippocampome.Org,¹⁰ our main neuroinformatics platforms, running.

The above considerations, however, do not exclusively apply to education and training, but can instead be extended more broadly to research grants for active, independent scientists. In the most general sense, once lots of brain data are up in the cloud, we gain tremendous efficiency for any fixed investment by identifying the most intelligent, passionate, creative, and hard-working among the scientists and trainees interested in studying those data, independent of their nationality or residency. This larger scientific community stands the best chance to explain what those data actually mean.

I realize that the timing of this Editorial may be considered at odds with the current era of immigration bans, border wall rhetoric, Brexit enactment, populism resurgence, and other nationalistic tendencies throughout the World. Yet governments and political fashions come and go at a faster rate than the accumulation of data necessary to reverse-engineer the brain.¹¹ It is ultimately up to the researchers in the field to forge and refine proper ideas for guiding progress. The right time is now.

⁶ Ascoli, G.A. (2013). Global neuroscience: distributing the management of brain knowledge worldwide. *Neuroinformatics*, 11(1), 1–3.

⁷ <https://www.nsf.gov/pubs/2016/nsf16588/nsf16588.htm>

⁸ <https://researchtraining.nih.gov/programs/fellowships/f31>

⁹ Ascoli, G.A., Maraver, P., Nanda, S., Polavaram, S., Armañanzas, R. (2017). Win-win data sharing in neuroscience. *Nat Methods*, 14(2), 112–116.

¹⁰ Wheeler, D.W., White, C.M., Rees, C.L., Komendantov, A.O., Hamilton, D.J., Ascoli, G.A. (2015). *Hippocampome.org*: a knowledge base of neuron types in the rodent hippocampus. *eLife*. 4.

¹¹ Roysam, B., Shain, W., Ascoli, G.A. (2009). The central role of neuroinformatics in the National Academy of Engineering's grandest challenge: reverse engineer the brain. *Neuroinformatics*, 7(1), 1–5.