

Consorting with big science

Large-scale consortia have benefited many fields chasing big data. In the face of new projects, neuroscientists should consider banding together.

New neuroscience initiatives across the globe have set their sights on big data. Private and public sectors are providing funding and resources toward collecting large-scale, largely hypothesis-free data. Although the ultimate utility of big data is often debated, the strategy for collecting and analyzing it also needs critical evaluation. The current mode of conducting science in small independent research groups may not be the optimal approach for amassing large, unbiased and shared data sets. To truly harness the benefits of big data, neuroscientists may need to band together to form more consortia than they currently do.

As neuroscience embarks on large-scale projects, the field needs to consider the practical power of forming consortia: large, formalized research collaborations often involving dozens of labs and hundreds of investigators across the globe. Given the strain on financial budgets available to support neuroscience research, the organizational structure of consortia may eliminate redundant and wasteful efforts across groups addressing the same research goal as well as provide the manpower needed to collect and analyze, in a standardized and rigorous manner, large data sets.

The benefit of consortium science is most apparent in the field of genomics (although it has advanced many areas of biology and physics as well). Faced with failures of replication and spurious findings, psychiatric geneticists formed the Psychiatric Genomics Consortium, which conducts mega-analyses of genome-wide data. Recently, they reported over a hundred regions across the genome associated with schizophrenia, which was made possible by including over 150,000 subjects¹. Amassing such a large sample size is simply beyond the resources of single research groups, which are not in a position to recruit the necessary numbers of subjects needed to conduct a well-powered genome-wide association study. The Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) Consortium was formed with this in mind and seeks to detect genetic variants influencing a range of brain traits. The stark reality is that some scientific questions cannot be tackled by a single laboratory.

Whether the aims of other big neuroscience projects such as the US government's BRAIN Initiative are best tackled by a consortium will depend on the scale and scope of the questions being asked. On the one hand, genomics deals with the narrow scope of genetic data, albeit on a vast scale. On the other, slated neuroscience projects aim to collect a wide scope of data encompassing different types of functional and structural data at varying degrees of resolution. One hope is to integrate these data, from the molecular to the behavioral, into a holistic understanding of brain function. The experimental, analytical and theoretical breadth required for this cannot be handled by one group. Even in a single domain, such as mapping the complete connectome of an organism, the sheer scale of the effort required is outside the

capabilities of a single laboratory. Although imaging technologies have improved, tracing the different cellular elements across tissue sections presents a major bottleneck in connectomics that has not been adequately automated and requires laborious human intervention. Reconstructing the entire brain of a single experimental organism presents a unique set of logistical challenges that a formal consortium may be able to help address.

In addition to the challenges of scaling up, there is also the issue of quality control and standardization. If data are to be shared across groups, there need to be agreed-upon methods for its collection and analysis. Just as genomics consortia have established standardized analysis pipelines for calling genetic variants, neuroscience needs to establish basic protocols that allow comparison and integration of data collected across different sites and conditions. However, it is important to keep in mind that no amount of processing or analysis will make data meaningful if they are not collected under the proper experimental conditions. Setting these community standards for the field is outside the purview of a single group, and it will be necessary to create consortia to help address this. In essence, consortium discussion will allow a formalized forum for small groups to come together in large numbers to reach a community consensus.

Undoubtedly, forming consortia comes with risks and challenges. Academic career advancement is based largely on productivity as measured through research publications, and there's the danger that the contributions of single investigators will be lost in the sea of researchers. As such, incentivizing data generation itself will require reform at the level of institutions and funding agencies so that all contributors can be appropriately acknowledged. Moreover, a large organization can be slow to adapt to changing research environments and can lack the flexibility of smaller groups. There is also the danger that fashionable ideas will dominate a particular research paradigm and hamper alternative approaches to a particular problem.

Will the rise in big science eclipse the need for smaller scale science? Consortia are not meant to supplant all current scientific efforts. There have been and continue to be important contributions made by more focused and smaller scale investigations. Not all scientific questions necessitate forming a consortium. Consortia are best used when they produce a data resource that will be useful to the larger community, such as the Human Genome Project or Allen Brain Atlas. Such projects are inherently hypothesis free and their main goal is data output. This is often a critique of what is pejoratively referred to as industrial or factory science. Yet this output can form the bedrock for more targeted and hypothesis-driven research carried out on a smaller scale. ■

1. Schizophrenia Working Group of the Psychiatric Genomics Consortium. *Nature* **511**, 421–427 (2014).