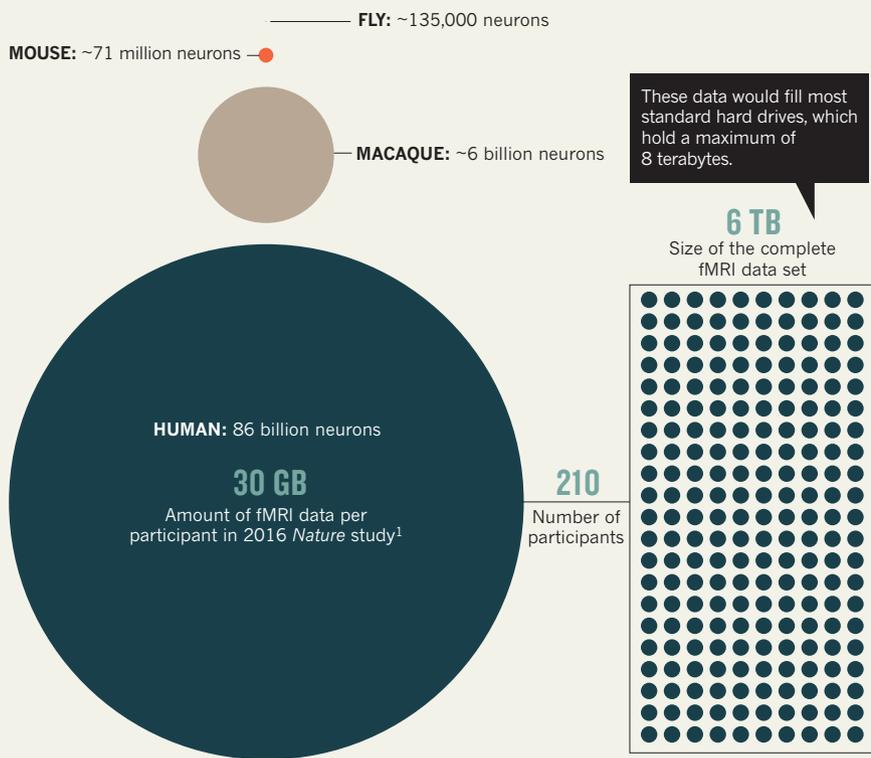


## BIG DATA BY THE NUMBERS

Mapping the brain presents an enticing challenge — and a computationally daunting one. Here's how many data one study last year generated.



the project at the Washington University School of Medicine in St. Louis, Missouri.

Few smaller-scale projects release their data, however — possibly because they don't have to. A few journals require all data supporting published findings to be made available to the community, but by and large, data sharing is not incentivized. There is “no strong impetus” to do that bit of extra work, says Grayson.

The conventional academic model doesn't help. Researchers typically develop hypotheses and work on their own ideas independently of peers in their group. In such an environment, research does not drive people together — it pulls them apart, says Hongkui Zeng of the Allen Institute for Brain Science in Seattle, Washington. “You need to distinguish yourself. To establish your identity in the field, you have to do something different from others.”

Zeng joined the Allen Institute in 2006 in search of a culture change: the institute sets out ambitious five-year goals that require teams to work collaboratively and systematically, driving a project to completion rather than piecemeal, as can happen in individual labs.

When it comes to the brain, ‘complete’ can be a moving target. But so, too, is the neuroscience toolset. During his Society for Neuroscience talk, Chiang lamented that it's taken ten years to map half the fly brain. Working with physicists at Taiwan's Academia Sinica, Chiang's team has started to use a technique called synchrotron X-ray tomography to boost data-acquisition speed dramatically.

“It took less than 10 minutes to image a fly brain containing thousands of Golgi-stained single neurons,” says Chiang, whose crew is now trying the method in mice and pigs. They plan to integrate confocal and X-ray images on a single platform from which scientists can download data. “With synchrotron X-ray imaging, mapping the human connectome at single-neuron resolution is now more realistic,” Chiang says. How easy it will be to meld the maps with other data remains to be seen. ■

**Esther Landhuis** is a freelance science writer in the San Francisco Bay Area of California.

1. Glasser, M. F. *et al. Nature* **536**, 171–178 (2016).
2. Biswal, B. B. *et al. Proc. Natl Acad. Sci. USA* **107**, 4734–4739 (2010).
3. Tomasi, D. & Volkow, N. D. *Proc. Natl Acad. Sci. USA* **107**, 9885–9890 (2010).

### CORRECTION

The Technology Feature ‘Metabolomics: Small molecules, single cells’ (*Nature* **540**, 153–155; 2016) erroneously stated that Matthias Heinemann was a former postdoc in Renato Zenobi's lab. Although he worked with Zenobi, Heinemann was a postdoc in another lab at the time. Also, Heinemann's background was in biochemical engineering, not analytical chemistry.

their data”, says Russell Poldrack, a psychologist at Stanford University in California who uses neuroimaging to study learning and memory. It could be “a generational thing”, he says: millennials are “much more into sharing code and data than my generation”. Poldrack worries that top minds might leave the field out of frustration with science “not aligning with the values they think it should have”.

But slowly, attitudes are shifting — first those towards software, then data. Conventionally, neuroimaging labs spend a lot of time downloading and installing the same beta software, “hacking through various software malfunctions and computing bottlenecks, writing redundant chunks of code and implementing their own data-management solutions to tackle the same problems”, says David Grayson, a neuroscience PhD student at the University of California, Davis. Worse, many non-research tasks are relegated to students, postdocs and young investigators, who tend to be tech-savvy, but “did not sign up to be sys-admins”, Grayson says.

The International Neuroinformatics Coordinating Facility (INCF), a non-profit organization based in Stockholm, was created in 2005 to develop and promote standards, tools and infrastructure for brain researchers around the globe. A few years later, the United States launched the Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC) as a

platform for sharing neuroimaging computational tools. Back then “no one was even thinking about sharing data, only software”, says Nina Preuss, a programme manager for the NITRC, headquartered in Washington DC.

That changed in late 2009, when researchers at the Nathan S. Kline Institute for Psychiatric Research in Orangeburg, New York, released resting-state functional MRI (fMRI) data into the NITRC from more than 1,200 volunteers, collected for the 1000 Functional Connectomes Project (FCP). These were just pooled raw data — yet within a few weeks, NITRC users had downloaded the data set 700 times. “There was such a pent-up demand for data people could freely download and play with,” says Preuss.

Download numbers soared to the thousands once the authors had cleaned up the fMRI data and made them searchable. After the data were published<sup>2</sup>, the paper logged more than 1,000 downloads in the first 2 weeks. In the same year, the first paper by independent authors — who had downloaded the consortium's fMRI data for their own analyses, but weren't involved in collecting it — was also published<sup>3</sup>.

Since the HCP made its first data set available in March 2013, dozens of outside labs have published papers analysing the project's data. In total, the HCP has released some 50 terabytes of brain-imaging data on more than 1,000 people, says Jennifer Elam, an outreach coordinator for