OPEN SCIENCE

Shared science's time to shine

Sharing research tools and data with other scientists brings many benefits, such as using fewer animals, improving reproducibility and increasing study sample size, but the practice still needs to be more widely adopted.

Charlotte Harrison

Three mice explore their surroundings, sometimes bumping into and blocking each other, their behavior captured on video. As they bustle around, software overlays dots on key body parts of each mouse, the dots connected by colored lines to track the animal's position and movement. The software doing the tracking is called DeepLabCut, a deep-learning method for determining the pose — or geometrical configuration — of animals in a non-invasive way[1,2]. Developed through a collaboration of university scientists, what's notable about the software is that it is broadly and freely shared with the research community.

The decision to openly share the software and underlying code was a natural way to push science forward by letting the research community build on and contribute to the tool, notes Alexander Mathis, a computational neuroscientist at the École Polytechnique Fédérale de Lausanne, Switzerland, and one of the team behind DeepLabCut. Mathis says, "To build on what others have made, you don't start from scratch, and you involve other researchers." Indeed, DeepLabCut builds on many open-source software tools and more than 100 people have contributed to the source code via the software-development website GitHub. This means, for example, that DeepLabCut can now be tailored to common behavioral experiments, such as multi-animal studies. And that's something that couldn't be done with a commercial product because you probably wouldn't have access to the code, he notes.

Many studies in the biomedical sciences have used the software, including those that incorporate it into other models, such as a computer-based method to detect bone destruction in people with arthritis[3], as well as for gait analysis of mice after stroke[4], cardiac physiology assessment in zebrafish[5] and behavior classification in lab-housed birds[6]. Its reach extends further, having been used to study wildlife and pets, and has more than 1,000 citations "I'm excited about open-source science…because I think it's the best way to make progress," Mathis says.

Connecting the world by sharing research tools and data. Credit: NicoElNino / Alamy Stock Photo

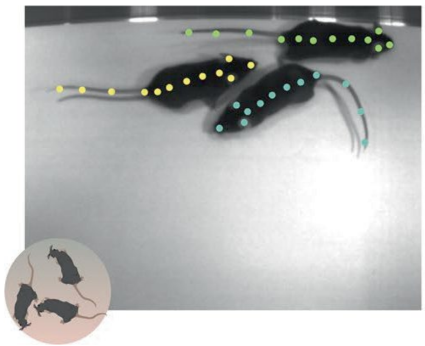
Like Mathis, other researchers and labs have devised open-science platforms and initiatives in animal research that are freely available to the community. The motivations behind, and experiences of, these initiatives are varied but often overlapping, and include accelerating research, reducing animal numbers and costs, boosting collaboration, and improving the quality and reproducibility of data.

The need to boost reproducibility was what motivated Marco Prado, a biochemist-turned-neuroscientist at the University of Western Ontario, Canada, to create MouseBytes, an open-access repository of cognitive testing data.

When Prado moved into neuroscience to study behavior in mouse models of Alzheimer's disease, he was struck by the variability of data in the literature and how hard it was to reproduce. Then, around 12 years ago, he learned about touchscreens as a technology for cognition testing. In touchscreen experiments, for example, a mouse is placed in a chamber and responds to images on the screen by poking its nose at the one it hasn't seen before. If its choice is correct, the mouse is rewarded with strawberry milkshake. "We were amazed to see how easy we could reproduce published data of mouse models of Alzheimer's," he says.

Touchscreens enabled Prado and colleagues to show that attention deficits in a mouse model of Alzheimer's disease (3xTG-AD mice) were seen in several labs even when a different genetic background was used[7]. "At the same time, we realized that this technology lends itself to data sharing really well because everything is automated and you generate datasets that are standardized," he says. Moreover, using computer-based analysis to measure animal behavior removes human subjectiveness from data analysis and interpretation.

MouseBytes enables researchers to store, share, visualize and analyze cognitive data. An enhanced version, MouseBytes+, allows



Identification and tracking of mice with DeepLabCut. Adapted from ref. [2], under a Creative Commons licence CC BY 4.0.

users to integrate MouseBytes data with other data such as those from imaging and photometry[8]. The platform is compatible with commercial touchscreens or those derived from open-source methods, and its sister website, touchscreencognition.org, is used for knowledge dissemination.

MouseBytes is a work in progress, but there are now data from more than 3,000 individual mice in the database. Prado hopes that soon other researchers will conduct metadata analyses, answering questions about the roles of different strains or sexes of mice in the touchscreen tasks. "People will be able to download these data and answer questions without doing any experiments," he says.

Mouse interacting with a touchscreen during a cognitive test. Adapted from ref. [17], under a Creative Commons licence CC BY 4.0.

Fewer experiments, more subjects

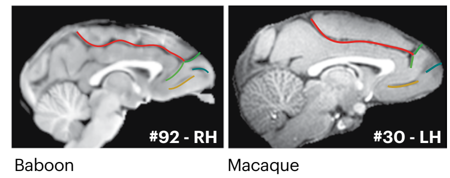
A sharing platform that is beginning to bear fruit in this regard, generating findings based solely or partly on existing datasets, is the PRIMatE Data and Resource Exchange (PRIME-DRE). This resource contains the imaging data (functional, diffusion and morphometric MRIs) of more than 800 non-human primates from 22 organizations[9,10]. Data from PRIME-DRE combined with data from other existing data sources enabled University of Lyon researchers to show that an area of the frontal cortex linked to higher cognitive functions previously thought to only exist in the human brain was actually present in other species and evolutionarily conserved[11]. The sample size in that study (197 humans, 225 chimpanzees, 88 baboons and 80 rhesus monkeys) was much higher than those in traditional non-human primate studies[11].

Non-human primate studies have typically used the smallest possible number of animals, both for ethical and cost reasons. Hence, there is a unique need for sharing data in non-human primate imaging, highlights Charles Schroeder, a Senior Research Scientist at the Nathan Kline Institute in New York and one of the team behind PRIME-DRE. "It's very expensive, it's highly regulated, and so it's becoming a very limited resource," he says.

PRIME-DRE was inspired by the success of data sharing with the Human Connectome Project as a way to create/build bigger, information-rich datasets and to promote collaboration. Sharing data makes it possible for people to put together large amounts of information in a meaningful way, meaning you "get more bang for your buck for each non-human primate study that is conducted," says Schroeder. Registered users of PRIME-DRE have full access to datasets, and the right to use them non-commercially. More than 30 papers have been published so far. A recent paper used PRIME-DRE data to complement human data; new findings about the function of the motor cortex showed that this brain area is a composite of two distinct and spatially interleaved systems, only one of which controls bodily movements[12].

Funding concerns

Giorgio Gilestro, an open-science advocate from Imperial College London studies sleep in fruit flies. A lack of suitable commercially available tools for measuring sleep and inactivity in this model led him to build his own hardware–software tool known as an ethoscope. The ethoscope facilitates real-time, high-throughput analysis of fly sleep behavior. During experiments, individual flies in test-tube-sized chambers are given a puff of smelly air – such as the odor of fermenting fruit – and then monitored by video to see if they wake up. The tool enables Gilestro's lab to process about 4,000 flies a day. They showed that



Non-human primate brain MRI. Adapted from ref. [11], under a Creative Commons licence CC BY 4.0.

flies, like humans, process complex sensory information when they are asleep, and that their responses are modulated by their internal state[13].

Every tool or software he and his lab create is shared with the research community on sites such as GitHub, Bookstack and Notion before it is published in the traditional journal format. But funding for hardware is problematic, particularly as hardware is not recognized by many funders, such as the Chan–Zuckerberg Initiative, that fund software. This sentiment is echoed by Alik Widge, from the Translational NeuroEngineering Laboratory at the University of Minnesota, USA, who with colleagues recently developed an open-source controller for operant training that facilitates data reproducibility and sharing[14]. He notes that obtaining funding specifically for hardware projects is an extreme challenge because you need a very large demonstrated user base. Gilestro advocates that open-source hardware can arguably make a bigger difference to the field than software, as it opens up the possibilities of what types of experiments researchers can do.

The difficulty to get the funding needed to build platforms, share resources and data, promote data reuse and ensure their long-term sustainability is an obstacle noted by several researchers. Without it, promising projects can fall by the wayside. For example, a UK-based platform of tissue, resources and information derived from animal models of breast cancer, dubbed SEARCHBreast, closed in 2018 owing to a lack of funding, and plans for similar initiatives for other diseases faded.

Large, often multidisciplinary teams are needed to get projects up and running. MouseBytes took "millions and millions of dollars" according to Prado; "you don't do this type of work with your normal grants," he says. Funders also have a role to play in incentivizing researchers to share; many funders require data to be made available upon publication. For example, the US National Institutes of Health (NIH) in January of this year mandated that all

Box 1 | Sharing data

The FAIR Data Principles state that findings and data behind research outcomes are findable, accessible, interoperable and reusable. To ensure data are both findable and reusable, the data need to be put in the best possible place so that they have the highest chance of being re-used effectively. This is a problem in animal research, which has diverse datasets, meaning animal research could be scattered all over the place. Many platforms allocate a digital object identifier (DOI) to information so that it is traceable. But finding data or other specific information without any prior knowledge can be challenging, and points to the need for better data-retrieval strategies. Some key repositories are listed below.

US NIH-supported repositories: https://sharing.nih.gov/data-management-and-sharing-policy/sharing-scientific-data/repositories-for-sharing-scientific-data

The Open Science Framework repository; links several platforms: https://osf.io

OpenAire; infrastructure to support communication infrastructure for European research: https://www.openaire.eu

The journal *Scientific Data* has a list of repositories organized by subject area: https://www.nature.com/sdata/policies/repositories#neurosci

NIH-funded projects share resultant data. Plenty of platforms are now available to share data (Box 1), protocols, software and other research material.

Adoption issues

But despite the plethora of platforms "we still have a problem of adoption of sharing initiatives," says Céline Heinl, a scientist at the German Centre for the Protection of Laboratory Animals. The issues holding back sharing are multifaceted. A culture of fast timelines, increase in workload (especially if protocols are shared or registered prior to studies), and the need for community standards within diverse animal research all get in the way.

Heinl and her team developed a tool for pre-registration of animal studies, animalstudyregistry.org to prevent selective reporting. This is a problematic area in animal research that could impede translation into human research[15]. When talking to many researchers about pre-registration, Heinl found that they didn't know about, or use, all the tools available for sharing and open science. "Most tools are adapted to animal research, so could be used already," says Heinl. To increase awareness of such tools, she and her colleagues created an Open Science Toolbox for Animal Research. She emphasizes that sharing should not be made compulsory, but researchers should be encouraged and incentivized to share so that they know why they are sharing.

The need to convince the community is also a problem faced by Prado and the MouseBytes team. Touchscreens are becoming a popular tool for mouse cognition testing worldwide, but currently around 70–80% of the data in MouseBytes comes from his lab or direct collaborators. "There is sort of an energy of activation that needs to be overcome before people in the behavior community see the importance of sharing," he says. His team is currently working to convince the community, working with philosophers of science and qualitative research specialists to try to understand what the barriers are. Prado hopes that in the future, animal research follows in the footsteps of the structural biology and genomics fields, where researchers routinely put their data in a go-to database – the Protein Data Bank for structural biology or the Gene Expression Omnibus for genomics. Why doesn't animal behavior have that?

The workload is one area of particular concern for already busy researchers. Matthew Grubb, a neuroscientist and data-sharing advocate from King's College London highlights the organizational effort needed to get data in a state that is ready to share. "That's surprisingly hard and could be a real stumbling block," he says. To share his lab's data, he needs to make sure that published data are backed up with the correct files, that they're cross-referenced to the right spreadsheet file, and that users will be able to follow the data because they are all in the right place. "It's a bit like doing a big sort out of your house. It sounds okay at first, but when you start you're like, oh my, where am I going to put this?"

Annemarie Lang developed AniMatch, a website for Europe-wide sharing of surplus tissue from small animals, while a PhD student at Charité-Universitätsmedizin Berlin, Germany. A vet by training with a keen interest in the 3Rs, the idea for AniMatch came about when a colleague asked her about sourcing tissue, and she realized that there was no central base for sharing tissue. Now a post-doc at the University of Pennsylvania, USA, the time taken to run AniMatch is challenging, especially as that time doesn't contribute to the traditional notions of academic career development. "It's nothing that I can put on my CV," she says, adding that she feels time needs to be primarily spent analyzing data or writing papers and grants if you want to become a professor. "You are not in a rewarding system, to be honest," she says, referring to the fact that volunteer activities are not taken into account in faculty applications. For now, she focuses on consultant-type activities, advising other researchers on tissue sharing.

Shining through

With everything said, might some people be a little bit anxious about sharing, especially for data or resources that they've worked so hard to generate? "I think the idea that if you put your data and so on out there, then you're suddenly going to get swarmed by parasitic people who get all the credit is a bit misplaced," says Grubb.

For Gilestro, a key positive of sharing is the more relaxed research environment. In his experience, the perceived fear of being 'scooped' can lead to self-inflicted stress and rushed experiments. "Being open-source allows us to be more creative at a slower pace and enjoy ourselves a bit more," he says.

The lower cost of open-source software, hardware and available data also means that they have a place in low-income countries, for example via the TReND in Africa program that promotes and facilitates the use of open-source biomedical resources for African researchers[16].

Researchers note that, based on other disciplines, early-career researchers are more open to the idea of sharing data and transparency and are not so afraid of sharing. Even established scientists, or those wanting to venture more than just depositing data or other resources shouldn't be put off about anything that seems unfamiliar — 3D printing or the programming language Python, for example — as they don't need to learn a new set of skills. "Open-science tools are so accessible, some of them are just really web-based that anyone with a little bit of good intention can learn," says Gilestro. "Don't be afraid of jumping into open science." ❐

Charlotte Harrison ✉
Science writer, Canterbury, UK.
✉*e-mail: charlotte.l.harrison@gmail.com*

Published online: 31 July 2023
https://doi.org/10.1038/s41684-023-01219-9

References

1. Mathis, A. et al. *Nat. Neurosci.* **21**, 1281–1289 (2018).
2. Lauer, J. et al. *Nat. Methods* **19**, 496–504 (2022).

3. Miyama, K. et al. *Arthritis Res. Ther.* **24**, 227 (2022).
4. Weber, R. Z. et al. *BMC Biol.* **20**, 232 (2022).
5. Suryanto, M. E. et al. *Biology* **11**, 1243 (2022).
6. Wittek, N. et al. *Behav. Res. Methods* **55**, 1624–1640 (2023).
7. Beraldo, F. H. et al. *elife* **8**, e49630 (2019).
8. Memar, S. et al. *Sci. Data* **10**, 210 (2023).
9. Messinger, A. et al. *Neuroimage* **226**, 117519 (2021).
10. Neff, E. P. *Lab Anim.* **49**, 139–143 (2020).
11. Amiez, C. et al. *Nat. Commun.* **10**, 3437 (2019).
12. Gordon, E. M. et al. *Nature* **617**, 351–359 (2023).
13. French, A. S. et al. *Nature* **598**, 479–482 (2021).
14. Dastin-van Rijn, E. M. et al. Preprint at *bioRxiv* https://doi.org/10.1101/2023.02.03.527033 (2023).
15. Baker, M. *Nature* https://doi.org/10.1038/nature.2015.19101 (2016).
16. Tsanni, A. *Nature* **582**, 138 (2020).
17. Nithianantharajah, J. et al. *Sci. Rep.* **5**, 14613 (2015).