Open access: Sharing your data is easier than you think

Below is the correspondence that I sent to Nature in May 2014; it was edited and subsequently appeared Nature 509, 33; 2014; `http://dx.doi.org/10.1038/510340c`).

In the ongoing debate over open data-sharing, two practical concerns regarding costs in implementing these policies [1] were raised that I think are unfounded.

Storing large-volumes of raw data can indeed be expensive, but in many cases the useful items to share are highly-processed and relatively small, well under a few gigabytes. e.g. the recently published mouse connectome is available as a three-megabyte Excel spreadsheet derived from many gigabytes of raw data [2]. Nor is there a shortage of repositories. Many institutional and international databases are freely available and well supported (e.g. zenodo.org is maintained by CERN). More repositories are likely to become available as we learn how best to share our data.

Sharing computer code may not require as much time investment as one thinks. The modest proposal [3] suggests that sharing code, irrespective of whether it is portable or documented, can be useful to others. Often many technical details are omitted from papers, and the code is the source of that information. People releasing code are under no obligation to maintain it, and the code should be seen as a valuable part of the scholarship relating to the paper. It is also often in your best interests to write code assuming that at some point you will share it; quite often it is reused by others in your lab (or even yourself in 6–12 months).

We have much to gain by freely sharing our data and code, and the sooner we do it, the better.

Stephen Eglen, Cambridge, UK.

[1] Practical costs of data sharing (Nature 509, 33; 2014) `http://dx.doi.org/10.1038/509033b`

[2] A mesoscale connectome of the mouse brain (Nature 508, 207–214; 2014) `http://dx.doi.org/10.1038/nature13186`

[3] The case for open computer programs (Nature 482, 485–488; 2012) `http://dx.doi.org/10.1038/nature10836`