My New Year's resolution for 2020 is to try harder at sharing data

Could 2020 be the year when biological anthropologists start to share their data routinely? Turner and Mulligan's (2019) article is an excellent blueprint for how, what, where and why to share data. Their case, developed in conjunction with a much wider group of scientists, is logical and straightforward to follow. We need to use it as a framework for aligning our discipline - study by study, article by article - with the movement for more open and transparent science, as well as shifting our academic culture to one of collaboration rather than competition.

The data sharing movement – and the debate that goes along with it – is not new. But it is now becoming mainstream, as journals and funders increasingly insist on researchers sharing data, or at least giving a reason for why they are not sharing them. Yet, I write this with an embarrassed blush. I agree wholeheartedly with data sharing. I have shared data and have benefited a great deal from data (and code) that others have shared generously with me. But, I have sometimes declined to share data in the past, and when asked to share data I have not always done so in a timely fashion.

In common with many people, I have been apprehensive about getting 'scooped', reluctant to reveal a raw dataset until it has been mined thoroughly. But that speaks to one of the major cultural shifts we need to make if data sharing is to become routine. Datasets are unlikely ever to be mined fully by one researcher or even a research team. Fresh eyes bring fresh perspectives, and different people are interested in different things. Several years ago, our lunchtime discussion group tackled data sharing. A colleague made a comment that I return to time and time again when thinking about data sharing. "The data are not the intellectual property" he said. "The ideas are. And what are the chances of someone having exactly the same idea as you when it comes to a dataset with so many variables, individuals and species?" Of course, there is convergence in research, especially as topics, theories and methods gain and decline in popularity over time. We've all been to a conference and thought "That's a cool idea – I'd like to do that." A quick internet search brings up multiple newsfeeds about scientists being beaten to a high impact publication by a rival research group, sometimes after sharing data. In a competitive academic world, these concerns have a legitimate basis, and we must discuss them openly, and address them, if data sharing is to become something we want to do rather than something we are compelled to. I sincerely hope the planned symposium at the AAPAs allows us to have this conversation and in the process reassure our community that sharing data is a positive act rather than one that is personally harmful.

Even if getting scooped is not a major concern – and who ever manages to investigate all the ideas they have – we must think about other practicalities surrounding data sharing if it is to succeed. In 2008, Andrea Cardini and I wrote an opinion piece about data sharing, 'Anthropology from the Desk', in the *Journal of Anthropological Sciences* (volume 86, pages 209-212). Several of the issues we highlighted, such as the need for suitable and accessible repositories and means of 'future proofing' access to research data, are well on the way to being resolved, as shown in Turner and Mulligan's (2019) article. A decade on, other issues still require more thought and may not be resolved swiftly. For open data to be useful, they must be intelligible without reference to the originator. The time and resources taken to collect data are often considerable, and many biological anthropology datasets stay in use for a long period. Thus, a major element for data sharing in our discipline is to encourage researchers to commit time to 'retrofit' datasets developed and collected in the era before data sharing. My tardiness in supplying data requested by another researcher often relates to mundane practicalities – data sheets that need to be collated or variables with names that need to be made comprehensible. If starting a project from scratch now, I would develop a data management protocol with sharing in mind, and take appropriate advice on that. Data curation is a skilled endeavour – at my institution we have an open access data specialist who can help throughout the design and implementation of a project. But, unfortunately, time travel to rectify previous sins of omission and commission is not part of the service. Another aspect of data sharing

Andrea and I highlighted, which also requires research time, can be summarised as 'caveat emptor'. All datasets, however carefully collected, have errors and need updating. Taxonomy shifts, for example. Most users are aware of this, and just as they would with a dataset they'd collected from scratch, will check for outliers and anomalies. Ideally, now that repositories are more accessible and stable, there will be a means for users of open access databases to report issues to the originator and have a clear audit trail of updates, corrections and amendments. This must be undertaken collaboratively and non-judgementally. And researchers, whether originator or user, will need to build time for such data management into their schedules. Over time, the benefits of this will outweigh the costs, but how we get there might be another interesting topic for discussion at the AAPAs.

In conclusion, Turner and Mulligan's (2019) article provides an incredibly solid foundation for data sharing within biological anthropology. It is now up to us, as members of that community, to act. I am fortunate to work on monkey morphology and evolution, and have always found the culture to be supportive rather than competitive. Sharing of ideas and resources has long been part of this and there has always been enough intellectual space for us all. This environment is due in no small part to the senior scholars – you know who you are - who trained and mentored the generation that is now training others. There are many other examples in biological anthropology, which is a relatively small community of researchers, with a strong and increasingly global set of research networks. We have lots of freedom to determine our research directions and approaches, and develop interesting new collaborations. This is an environment in which data sharing can thrive and make our discipline even richer. If the whole task at hand seems enormous, we should do our best incrementally at first, article by article. So, my New Year's resolution is that in 2020 I will try very hard to practice what I preach. I will take some time to make my datasets understandable to others and, when publishing, my default will be to share freely rather than say that 'data will be available on request'.

Sarah Elton

Durham University, Department of Anthropology, Dawson Building, South Road, Durham, DH1 3LE, UK.

Email: sarah.elton@durham.ac.uk