A post-doc's experience with open science

a case study of Dominique Roche

Key Points

- **Backup your data**
  Dominique now backs up his data after losing 6 weeks' worth of data in an accident.

- **Don't be afraid to share your data, even as a PhD student**
  Sharing your data can be scary, but identifying errors is important for science.

- **Prepare for how you're going to store and share your data**
  Thinking about how you're going to share your data ahead of time incentivizes you to be much more organized and careful in the way that you collect and store data.

About Dominique Roche

Dominique Roche is an aquatic ecologist at the University of Neuchâtel in Switzerland. His research focuses on species-environment interactions. He is presently an ambassador for Figshare and for the Center for Open Science.

It was during his PhD studying coral reef fish in the Great Barrier Reef at the Australian National University that he heard about data sharing.

" People were very passionate and tended to position themselves at opposite ends of the spectrum, arguing either strongly in favor of keeping data private or making it public."

While at a research station collecting data for his research, an unfortunate incident led him to lose a hard drive - as well as the backup - containing 6 weeks' worth of carefully collected data. These circumstances got him thinking about the need for not just a physical backup of his data, but more robust cloud-based backups.

Later that year, as part of a departmental reading group that meet weekly to discuss interesting journal articles, he and colleagues exchanged views about a paper titled "Data archiving in ecology and evolution: best practices" (TREE 26:61-65). The paper made a splash in the research community when it was published in 2011. Dominique found it interesting to witness the stark contrast in his colleague's opinions about data sharing and reuse.

Some researchers he knew to be very collaborative were surprisingly reluctant to share their data; they felt that the years worth of time and effort they had put into obtaining grants and painfully collecting data gave them a strong sense of ownership over them. In contrast, others felt that the scope and quality of their research was dramatically improved when they were allowed to access and reuse others' data.

As an empirical ecologist who generates primary data, Dominique understood his colleagues' concerns. However, he also clearly understood the benefits of promoting a stronger culture of sharing in science. Since he rarely re-used his own data after analyzing them for their intended purpose, Dominique was generally happy to share his data with other researchers. His only concern was with other people finding errors in his work which, for a PhD student, could be unsettling and potentially harm his reputation. Nevertheless, he recognized that "identifying errors is also important for advancing science, even if it can be uncomfortable."

Following this meeting at the journal club, Dominique thought carefully about his colleague's concerns, weighing the pros and cons of data sharing. Importantly, he wanted to make sure that everyone's voice was heard and that those who were keen to share their hard-earned data would be recognized for their contribution. As a result, he convinced his colleagues to come back to the table for discussions with the objective of reaching a consensus on the issue. These discussions led to a collaborative paper, spearheaded by Dominique, titled Troubleshooting Public Data Archiving: Suggestions to Increase Participation, now available in the Public Library of Science (PLoS Biology).

Dominique is now a post-doctoral researcher and regularly works with postgraduate students, encouraging them to consider the benefits of open science practices in their own research. In addition to mentoring students, he regularly gives talks about open science in Switzerland and abroad, and enjoys talking with his colleagues who are less familiar with data sharing practices.

"It's been interesting to see how people's mindsets have changed in the last few years as open science is becoming more mainstream," says Dominique. "Researchers at all stages of their career are more aware of the need to be more transparent and organized in their data management to increase the credibility of their research results."

Dominique has noticed that many senior researchers he works with are increasingly realizing the importance of training their students to carefully document their research from inception through to data collection and analysis. Researchers will often look back on data that was collected several years ago and attempt to combine it with fresh data to conduct new, exciting analyses. However, they often find that the old data was poorly documented and spend a lot of time trying to make sense of spreadsheets, sometimes unsuccessfully. "People who go through this process quickly understand how important it is to document data not just for others, but also for themselves," says Dominique.

Dominique has organized several training workshops on data management and open science. For the most part, he feels that young researchers are keen to see collaboration and transparency become the norm in research.

" They're interested in changing the motive of research from one that's focused on individual career advancement to one that's aimed at moving science forward."

Many journals in ecology and evolution are now requiring that authors make their data publicly accessible. For researchers who are wary about openly sharing their data - especially those who plan to carry out further analyses - Dominique suggests not shying away from journals that require open data.

Authors can safely archive their datasets on an online repository and simply opt for an embargo that keeps the data private until the authors are ready to release them. Basic one-year embargoes are available from most mainstream repositories but it is also possible to request extensions for up to 10 years.

"Thinking about sharing ahead of time incentivizes you to be much more organized and careful in the way that you collect and store data," says Dominique. "This is obviously good for science as a whole but it's also beneficial for researchers – in the end, being organized saves you time and often makes your life a lot easier!"
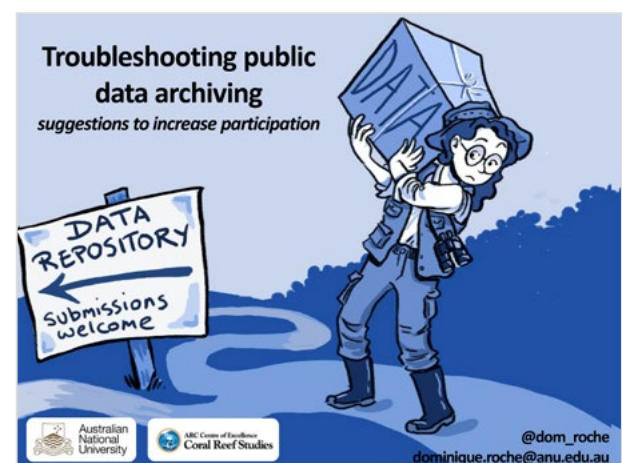
This lead Dominique to conduct a study on the reusability of data in his field, titled Public Data Archiving in Ecology and Evolution: How Well Are We Doing?. Dominique and his colleague carefully screened 100 papers in ecology and evolution journals that mandate data archiving.

They discovered that less than half of these datasets are complete; there were often missing variables or even large parts of the dataset. They also discovered that 65% of these datasets were archived in a way that made reuse difficult or impossible.

"In some sense, it's encouraging to see that many of these datasets are in a great shape," says Dominique. "However, I think it's clear from these results that open data policies are not 100% effective. Three things need to happen for data sharing to improve: journals need to clarify and communicate their policies better, universities need to educate researchers on how to share their data effectively, and funding agencies need to recognize data sharing as an important scientific contribution to incentivize this behavior."

Since first becoming familiar with data sharing back in his PhD, Dominique began storing his data in the cloud and sharing it publicly. He uses Figshare to do this and finds the ability to generate a private link to share with journal reviewers prior to publication particularly useful. Links to all of his published datasets since 2013 are listed on his personal website.

Alongside his work in aquatic ecology, Dominique is actively pursuing research on data sharing in the hopes that it will both speed up and improve the reproducibility of scientific research.

#G2B2014 talk:
Troubleshooting public data archiving: suggestions to increase participation. Illustration by Ainsley Seago (@americanbeetles).

See Dominique's research on figshare!