Building workflows for data publication and reuse

Patrick Charbonneau

Professor of Physics, Duke University

Recommendations

- Start with simple steps for data sharing and expand your workflows over time
- Share data in a way that facilitates reuse, allowing others and yourself to pursue future exploration, in ways perhaps unanticipated when the dataset was created
- Incorporate data in institutional evaluation processes, as recognition of responsible, reproducible, and ethical scientific practices

During the early years of his independent career, Patrick grew increasingly aware that the data generated through his research were hard to recover in case they were needed for further internal validation, or for sharing with collaborators and other scientists. Motivated by an interest in ensuring persistence of the data, and accountability towards data stewardship over time, Patrick and his team created processes to regularly deposit in a repository the data produced as part of their research; for the last decade, Patrick's group has shared the data underlying each of their articles.

With reproducibility as a focus, Patrick and his group started by sharing the data points underlying each of the figures in their papers. Over time, they expanded their workflows to proactively think about data sharing at earlier stages of data collection, and to share a broader range of data, scripts, and codes generated throughout the research process. Patrick's group worked with the library as they developed the flow of their data, the protocol for exchange, and approaches for public data deposition.

Patrick highlights that making data available is straightforward in practice, and can enable publications to have broader impact. He reports datasets he produces and shares in his annual reviews, and would like data contributions to become part of regular assessment of faculty, as a signal of good and responsible research practice and demonstration of productivity. Data contributions often result in new archives and resources for the community, and this should also be recognized in evaluation.