Reproducibility and Open Science with ReproZip

Vicky Steeves (1) and Remi Rampin (2)

(1) New York University, Center for Data Science & Division of Libraries, United States, (2) New York University, Tandon School of Eningeering, United States

Reproducibility is a core component in open research. While it's imperative that scientists share their code and data, this does not necessarily assure that research is verifiable and reproducible. To make the most use of our own and others science, the ultimate goal should be to make our research open and reproducible.

To help researchers create open, reproducible research, we have created ReproZip (https://reprozip.org). ReproZip is an open source tool that creates small, self-contained reproducible packages in as few as two steps. ReproZip packs everything necessary to reproduce research, including data, code, libraries, configuration, and even parts of the computational environment. These small packages can then be posted online, shared with collaborators, or sent in as supplementary materials with publications. Readers can then unpack these files and reproduce the research, even across different operating systems!

By creating easy, open source infrastructure, it is our goal to make reproducibility easy for everyone, from graduate students to senior faculty members. In our talk, we will demonstrate how to integrate ReproZip into the research workflow through looking at use cases in scientific publication, research archiving, and open science sharing, with a focus on geoscience. Current use cases are indexed at https://examples.reprozip.org, and the list is constantly expanding.