

EarthCube Research Coordination Network "What About Model Data?" - Best Practices for Preservation and Replicability

1. Project Overview

There is strong agreement across the sciences that replicable workflows are needed for computational modeling. However, recent efforts to standardize data sharing and preservation guidelines within research institutions, professional societies, and academic publishers make clear that the scientific community does not know what to do about data produced as output from computational models. The massive size of the simulation outputs, as well as the large computational cost to produce these outputs, makes this not only a problem of replicability, but also a "big data" problem. While most researchers that produce simulation output would welcome more use of their output products, and many end users would welcome more data availability, the reality is that we are producing far more simulation output than can be reasonably stored in repositories. The goal of this Research Coordination Network (RCN) project is to provide simulation data management best practices to the community, including publishers and funding agencies.

2. Data Archiving Rubric

The first product of this RCN project is a rubric to be used to assist a researcher in determining what simulation outputs should be deposited in a FAIR aligned community repository to communicate knowledge. The rubric is built to help researchers make decisions about what simulation output needs to be shared via a repository, i.e. made accessible and preserved a sufficient time to satisfy the requirements of publishers and funding agencies. The rubric is a list of simulation/experiment descriptors, organized into themes. The rubric guides researchers through these descriptors to develop a recommendation on whether the researcher should preserve few simulation outputs, some outputs, or the majority of of the simulation outputs

Although questions about what outputs should be preserved are often broadly dictated by the research proposal and the expected user community, the rubric and use case examples are useful for making decisions on specific components of a project should be preserved (e.g. input data, pre and post processing codes, raw model output, and processed output), and for communicating with end users about why not <u>all</u> of these components can be made available.

3. For More Information

To find the project's data archiving rubric and guidance documents, please visit the project website: <u>https://modeldatarcn.github.io/</u>. To join the project's mailing list or to express your interest in future project workshops and outcomes, please contact the project PIs:

- Gretchen Mullendore, gretchen.mullendore@und.edu
- Matt Mayernik, <u>mayernik@ucar.edu</u>
- Doug Schuster, <u>schuster@ucar.edu</u>