Data Management Plan

Types of data produced

The primary data types produced by this project will be computer code and model simulation results. Modeling will be conducted using MATLAB software, so code will be generated in .m format and model output will be generated in .mat format and then converted to NETCDF format for archiving and sharing.

Quantity of data produced

Each 250-year simulation with our global methane cycle model will generate approximately 1GB of output, saved at annual timesteps. Many hundreds of model simulations will conducted during the development and testing of our model, and therefore this project will generate will generate hundreds of GB of data per year. The majority of this data will not be of interest to broader scientific community, since it is only generated for testing purposed. Model output from finalized model versions (those used in publications) will likely total ~10GB per year.

Code availability

All model code developed in this project (including preliminary testing code) will be preserved on the University of Rochester Bluehive computer cluster, and made available upon request. Code for finalized model versions (those used in publications) will be archived in a publicly available Zenodo repository, where they can be freely downloaded and used by the scientific community and general public (although the codes will require proprietary MATLAB software to run). Archived codes will be well documented and accompanied by brief user guides to ensure they are accessible, and contact information for Dr. Weber will be provided for troubleshooting.

Data availability

All model output produced by this project (including preliminary testing output) will be preserved on the University of Rochester Bluehive computer cluster in .mat format, and made available upon request. Output from finalized model versions (those used in publications) will be archived in the publicly available Figshare and BCO-DMO repositories in NETCDF format. Using both repositories will ensure that our data is as accessible as possible – Figshare will provide a convenient way to gain fast access to the data directly from our papers, whereas BCO-DMO will allow researchers to find our dataset through keyword searches. In both cases, data will be archived upon publication or within two years of generation, whichever comes first, and will be accompanied by README files that provide context for the model output.

Metadata

The primary metadata required for interpreting the model output is the geographical grid information, and model inputs (such as atmospheric methane forcing data). In all cases, the grid information and input files will be made available with the model output in the Figshare and BCO-DMO archives.

Data longevity and protection

Archived model output and computer code will be made available as long as the Zenodo, Figshare, and BCO-DMO archives allow. Output and code will also be preserved indefinitely on the University of Rochester Bluehive Cluster, where it is also backed up in a separate location. If data becomes unavailable in online archives for any reason, it can always be requested from the Principle Investigator who will transfer it from the Bluehive Cluster.