

Title: Help Wanted: Defining an Extensible, Understandable, and General (?) Naming Convention for Ecological and Environmental Data

Anthony Walker^{1*}, Martin G. De Kauwe², Paul Hanson¹, Belinda E. Medlyn³, Richard Norby¹, Sönke Zaehle⁴

¹Environmental Sciences Division and Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN;

²University New South Wales, Sydney, Australia;

³Western Sydney University, Australia;

⁴Max Planck Institute for Biogeochemistry, Jena, Germany.

Contact: (walkerap@ornl.gov)

Project Lead Principle Investigator (PI): Anthony Walker

BER Program: TES

Project: FACE Model Data Synthesis project

Project Website: <https://facedata.ornl.gov/facemds/>

Project Abstract: Your help is needed! Your feedback helping to define an extensible and understandable variable naming convention for ecological and environmental data would be greatly appreciated. The FACE Model Data Synthesis (FACE-MDS) project has been working for close to the past 10 years to synthesize data from FACE experiments and an ensemble of terrestrial ecosystem models to help improve our predictive understanding of the future terrestrial carbon sink. A critical element contributing to the success of the project has been a well defined protocol and QA/QC of model results. Integral to the protocol, QA/QC, and scientific analyses is the naming convention used to label the environmental and ecological data. A robust naming convention may not seem all that exciting but, agreed early in the project, the model output dataset with many variables and their associated naming convention allowed rapid assessment of model output. In many cases this assessment led to revised simulations and to the identification of bugs in the models' code. A large number of output variables also allowed in-depth assumption-centered model evaluation.

Many similar data synthesis and model inter-comparison projects exist. Thus an extensible, understandable, and possibly general naming convention would facilitate these kinds of projects. To achieve extensibility the naming convention must rely on a naming standard, that is hierarchical and ideally incorporates widely-accepted pre-existing scientific conventions. It also requires consideration of different classes of variables. We advocate for names that are readable and are broadly understandable without reference to a naming dictionary. As part of the FACE-MDS project we have been developing a naming convention with these goals in mind, but this is under development and community feedback will help to ensure many bases have been covered. At the highest level we categorise variables into mass stocks, mass fluxes, and traits (in the broadest sense). The next optional level is used for very commonly used variable names or abbreviations, e.g. GPP. The next level is to identify the location of the variable in the ecosystem, e.g. leaf. Further optional levels can be used to describe finer differentiation, e.g. species. Please add your two-cents, two-dollars, or preferred currency to the conversation.