Title: **ESS-DIVE Standards for 16S Amplicon Data Products, Comma-Separated Data Files and File-level Metadata**

Terri Velliquette[1], Pamela Weisenhorn[3], Les Hook[1], Jess Welch[2], Jack McNelis[1], Ranjeet Devarakonda[1], Susan Heinz[1], Michael Crow[1].

[1]Environmental Science Division, Oak Ridge National Laboratory, Oak Ridge, TN
[2]Information International Associates, Oak Ridge, TN
[3]Biosciences Division, Argonne National Laboratory, Argonne, IL

Project Abstract: We are researching and developing generic standards for 16S amplicon abundance tables, for the data stored in comma-separated values (CSV) text format files, and for minimal file-level metadata (FLMD) schema to help data users search and locate files archived in the ESS-DIVE.

The target for the first standard is to 1) identify a typical data structure for 16S abundance tables and 2) capture the most critical provenance in machine readable format to support data re-use in models and meta- analyses, all without overburdening the original data generator. After identification of common workflows, a set of samples were run through multiple workflows with different parameterizations to examine the effects on the generated 16S amplicon abundance table to identify which factors most strongly affected the comparability of output as well as the ease of generating abundance tables in a standardized data format.
The results of this cross-workflow comparison and draft standard for 16S amplicon abundance table and provenance metadata will be presented.

This work is being done in close consultation with Systems Biology Knowledgebase (KBase) project to ensure that 16S amplicon data products stored in the ESS-DIVE repository can be seamlessly transferred to and from KBase for analysis and with the National Microbiome Data Collaborative.

To build a community supported standard for CSV and FLMD, we researched existing standards and sought feedback from ESS investigators. In this poster, we present our work thus far with preliminary thoughts on our deliverables that also include a CSV format checker and FLMD extraction scripts.