DATA MANAGEMENT PLAN

The PI team takes seriously the need to share data among project personnel, and, following a period of exclusive use, the entire research community and the public. Sample and data access rules will follow standard protocols developed through extensive consultation with the research community, ICDP, and NSF, meeting requirements of the NSF Division of Earth Sciences Data policy (https://www.nsf.gov/geo/ear/2010EAR_data_policy_9_28_10.pdf).

Types of data, samples, collections, & other materials
This project will utilize lacustrine sediment and water-column samples. Any unused splits of sub-samples taken for this study will be archived in the sample repository of the Werne lab maintained in the Department of Geology and Environmental Science at the University of Pittsburgh. These sample splits will be available to other investigators upon request to the PI. The project is expected to produce Gbp of sequence data (fastq format), which will be stored on dedicated storage and backup drives. All data will be stored on off-network devices such as hard drives. All biogeochemical data and rates will also be stored online in Dropox or a similar platform.

Data Format
Data will be stored in formats generated by software utilized for a given analysis (e.g., as .dat files for isotopic analyses, fastq or fasta for next generation sequencing), as well as in Excel spreadsheets. Data about sample analyses will also stored in laboratory notebooks maintained by the Werne, Elliott, Newell, and Hamilton labs.

Data organization & quality assurance
Laboratory replicates and blanks will be collected according to standardized methods. All quality assurance and quality control efforts will be recorded in the dataset metadata. For isotope samples, the value of certified standards will be monitored throughout the run to ensure proper accuracy.

All geochemical, isotopic, and biological data reports will contain all critical information such as original sample IDs, run date/time, normalization data, sequencing strategy, and information on standard deviation of duplicate samples and check standards.

Data archiving & sharing
Data produced in this study will initially be stored electronically at the University of Pittsburgh on backed-up computer systems, and made available to the community upon request to the PIs. Within 2 years, data will be submitted to a national archive facility such as the Open Core Data (http://opencoredata.org, hosted at the NSF IEDA/Interdisciplinary Earth Data Alliance facility) and the NOAA Great Lakes Observing System (GLOS; glos.us) for long-term storage and preservation and for semantic enhancement, transformation for machine readability, and harvesting by relevant archives and domain repositories (e.g. NOAA Index to Marine and Lacustrine Geological Samples, Neotoma Paleoecology Database, MagIC/Magnetics Information Consortium, etc.). The project investigators will comply with the data management and dissemination policies described by NSF IEDA and NOAA GLOS.

Results will be published promptly in peer-reviewed journals such as *Geochimica et Cosmochimica Acta*, *Biogeosciences*, *Limnology & Oceanography*, *Earth and Planetary Science Letters*, *Quaternary Science Reviews*, *Organic Geochemistry*, and others. All data supporting these publications will be submitted as tables to be available in the supplemental online materials (unless incorporated into the main text of manuscripts).

Repositories for archived microbiological data and mechanisms for public access and distribution:

A long-term data sharing and preservation plan will be used to store and make publicly accessible the data beyond the life of the project. The data will be deposited into the Data Repository for the University of Minnesota (DRUM), http://hdl.handle.net/11299/166578. This University Libraries' hosted institutional data repository is an open access platform for dissemination and archiving of university research data. Date files in DRUM are written to an Isilon storage system with two copies, one local to each of the two geographically separated University of Minnesota Data Centers. The local Isilon cluster stores the data in such a way that the data can survive the loss of any two disks or any one node of the cluster. Within two hours of the initial write, data replication to the 2nd Isilon cluster commences. The 2nd cluster employs the same protections as the local cluster, and both verify with a checksum procedure that data has not altered on write. In addition, DRUM provides long-term preservation of digital data files for at least 10 years using services such as migration (limited format types), secure backup, bit-level checksums, and maintains a persistent DOIs for data sets, facilitating data citations. In accordance to DRUM policies, the (deidentified, if applicable) data will be accompanied by the appropriate documentation, metadata, and code to facilitate reuse and provide the potential for interoperability with similar data sets. Any publication, report, or completed data set for which we retain the copyright will be stored on DRUM. Supported formats include many common application file types. We will also use it for other reports such as undergraduate student theses, which may not be easily available to the public otherwise (graduate theses are made publicly available through the University of Minnesota Libraries).

All sequence data (fastq, fasta) produced in this project will be handled according to U.S. DOE Joint Genome Institute guidelines for data release (http://www.jgi.doe.gov/sequencing/collaborators/datarelease.html), including Genbank submission, web publication, approved data quality (http://www.jgi.doe.gov/sequencing/collaborators/finishing.html), and metadata standards recommended by the Genomics Standards Consortium. Genomic and metagenomic data will be made available upon publication or within 2 years of generation at the JGI websites including the Microbial Genomics portal (http://genome.jgi-psf.org/mic_home.html), the Integrated Microbial Genomes site (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi), and the Microbiomes site (http://img.jgi.doe.gov/cgi-bin/m/main.cgi) and through The National Center for Biotechnology Information (NCBI).

A long-term data sharing and preservation plan will be used to store and make publicly accessible the data beyond the life of the project. The data will be deposited into the Data Repository for the University of Minnesota (DRUM), http://hdl.handle.net/11299/166578 as described above. All biogeochemical data will also be uploaded to the NOAA Great Lakes Observing System (GLOS) for long-term storage and preservation of all data from this project. The project investigators will comply with the data management and dissemination policies described by NOAA GLOS.