Fall 9-19-2024

Effects of Biofilm Colonization on the Dynamics of Microplastics in Turbulent Flow

Liyuan Joanna Hou
Utah State University, liyuan.hou@usu.edu

Data Management Plan

This project will comply with the Division of Chemical, Bioengineering, Environmental, and Transport Systems (CBET) National Science Foundation policy on data management and sharing.

1. Expected Data Type

Analytical results, simulation models & data, images data, educational materials (e.g., handouts and fact sheets), meeting notes, progress reports, presentation slides, publications, and other materials will be produced in this project. Data from Hou's group include (1) analytical results including but not limited to Fourier-transform infrared spectroscopy spectra; (2) images of optical microscopes, atomic force microscopy (AFM), and scanning electron microscopy (SEM); and (3) microbial strains, biochemical assays and microbial growth data, and genomics sequencing data. Data from Berk's group include (1) 12-bit monochromatic images with a resolution of 1024 × 1024 pixels, to be recorded on a SanDisk portable SSD and backed-up on a Synology DS1812+ DiskStation; and (2) experimental data for dynamic trajectories of MPs with biofilms in excel sheet. Computational data from Su's group include (1) simulation settings including the simulation setup, the boundary conditions, and the modelling parameters; (2) output from the simulations including the settling/rising velocity profiles and the trajectories of the microplastic particles; (3) the raw/pre-processed raining dataset for the machine learning models; and (4) the obtained machine learning models. All computational data from Su's group will be stored as plain-text files and can be visualized using appropriate software, such as ParaView. Certain data may be presented in the form of plots, figures, and videos as appropriate. The projected total data size is estimated to be between 1 TB and 10 TB.

2. Data and Metadata Standards

Project electronic data are processed through workflows developed in the participating laboratories and may comprise commercial, open source, and in-house developed software and computational tools. Digital research data produced from our project includes raw data, associated metadata, data analysis software, and output files. All protocols, simulation reports and journal publications will be scanned and saved as PDF files. Original data and processed data in computers connected to analytical instruments and other text-based data will be stored in TXT, Excel, and Word files. Presentation and poster files will be stored in PowerPoint files. The biological sequence data will be stored in FASTA format. Images will be saved in uncompressed TIFF format. Time-series data from particle trajectories, as well as numeric data derived from the microplastic samples will be stored in Comma Separated Values (.csv) files. Metadata will be stored in Open Document Format (.odt) text files. Derived data includes maps of velocity and trajectories of particles, calculated by processing the images using in-house Matlab-based algorithms. This derived data will be stored in Comma Separated Values (.csv) files. For computational data, plain-text files will be used for human-readable data, and ParaView-compatible files will have no specific suffix as long as they are correctly formatted. Post-processing scripts will be written in a Unix shell or Matlab/Python format.

3. Policies for Data Access and Sharing

The management and stewardship of scientific data in this endeavor will strictly adhere to the principles of findability, accessibility, interoperability, and reusability (FAIR). The data produced within the proposed project will be disseminated through publication in rigorously peer-reviewed journals and through presentations at prominent national scientific conferences. Anticipated journal publications are expected to emerge both during the project and upon its conclusion. Once the submitted manuscripts are accepted for publication, the data utilized in these publications will be made fully and publicly accessible. Links to the corresponding data will be thoughtfully provided within the manuscripts themselves. As for biological

sequence data, it will be promptly uploaded to both the Repository of Algorithms, Instrumentation, and Data (RAID) as well as the esteemed National Center for Biotechnology Information (NCBI).

4. Re-use, Re-distribution, and Production of Derivatives

Certain project data may be retained until project completion to facilitate publication by PIs and students or to adhere to collaborating institutions' intellectual property policies, with potential non-disclosure agreements with industry partners. Data usage restrictions are minimal: (1) users must provide contact information and usage description, and (2) proper data source citation is mandatory. Intellectual property and rights for project-generated materials and data will follow policies of Utah State University (USU), National Science Foundation, and relevant entities. Open-access repository data can be used for academic research, education, and non-profit activities, with other uses requiring written PI authorization. Intellectual property from potentially patentable discoveries will be shared per Utah State University (USU) policies, subject to investigator and university agreements, and encompassing novel material compositions, mechanical properties, and unique combinations.

5. Plans for Archiving, Preservation, and Access to Data

The project will implement stringent data protection measures, rendering data machine-readable and well-defined, and will securely store it within computer systems equipped with backup mechanisms and the encrypted cloud sharing service, Box, at USU, ensuring data preservation in case of system failures or accidental deletion while supporting diverse file types and formats. Robust data archiving protocols will be maintained across all participating groups, with research data archived in the DigitalCommons@USU institutional repository, offering comprehensive file format support and providing persistent Uniform Resource Locators (URLs) and Digital Object Identifiers (DOIs) for datasets to enhance discoverability and citability. Preservation copies will be securely housed in Amazon Web Services, employing redundant storage across multiple facilities and undergoing regular integrity verification through checksums. The collected microplastic samples will be stored in the Experimental Fluid Dynamics Laboratory at the USU College of Engineering, while biological samples will be meticulously preserved at -80°C, complete with numerous replicate freezer stocks, within the environmental quality laboratory (EQL) at the Utah Water Research Laboratory.

6. Roles of Project Staff in the Data Management

Throughout the active phases of the project, including data collection, analysis, and publication, each Principal Investigator (PI) will serve as a dedicated data steward, meticulously documenting and managing the data. Dr. Hou, the designated PI, will oversee the execution of the Data Management Plan (DMP), coordinating with the project team and conducting quarterly reviews to ensure planned data deposition progresses as intended. Graduate research assistants, undergraduate researchers, and other project personnel will receive comprehensive data management training, sharing responsibility for adhering to the data-management plan. To facilitate the movement of samples and equipment between laboratories, a standardized nomenclature will be adopted, tracked in a shared spreadsheet. Detailed data history log encompassing file naming, organization, versioning, data entry, documentation, and change tracking will be established with regular six-month intervals. As the grant period concludes, the data will be systematically archived and shared according to the previously outlined procedures.