## PROJECT DATA ANALYSIS

**Types of Expected Data**

* Generated mass spectra data from urine samples.
* Types: Table, Chart, Graph, Figure, Raw, Processed, Filtered, Analyzed
* Existing datasets: preexisting metabolite databases
* Expected formats: .csv, .xlsx, .d, .mzML

# PRESERVATION AND STORAGE

This project is anticipated to generate terabytes of data and will require an external hard drive to store data throughout the project (5 terabytes in size). Project data will be stored here for the duration of the project, after which it will be stored indefinitely. Raw MS data will be uploaded to NIST for other researchers to use openly. Patient identification data will be stored in a secure location and password protected. Code will be stored in a GitHub repository (available to the public), and relevant metadata will be uploaded to FigShare. A detailed README file will accompany raw data and GitHub code.

NAMING CONVENTION

The .d files will be named based on the date the samples were run, day, month, and year (i.e. 24102024) with \_initial to indicate who ran it (i.e. 24102024\_LS) and a numerical identifier to indicate the sample it belonged to (i.e. 001\_24102024\_LS). All relevant sample metadata (i.e. identification) is stored in a .csv file named 2024DoA\_sample\_metadata.csv.

RELEVANT STANDARDS

* All instrumentation use is documented in logs that are instrument specific
* All laboratory work is documented in individual notebooks
* General good laboratory practices (GLP)
* Consistent file naming
* Data backups using 3-2-1 strategy.

DOCUMENTATION STRATEGIES

* All laboratory work will be documented in laboratory notebooks that will be kept in the lab. A detailed README will accompany the data processing workflow with the relevant code that is available.