Plan Overview

A Data Management Plan created using DMPTool

Title: Trace Element (TE) Shared Resource

Creator: Christian Lytle

Affiliation: Dartmouth College (dartmouth.edu)

Principal Investigator: Brian Jackson

Funder: National Institute of General Medical Sciences (nigms.nih.gov)

Funding opportunity number: PAR-19-301 Limited Competition: NIGMS National and Regional Resources

Grant: R24GM141194

Template: NIH-Default DMSP

Project abstract:

Data generated by the TE shared resource is the sole property of the client and are not openly shared by the TE shared resource. TE analyses are not considered human subjects research because personal identifiers or other individual-specific data is never supplied with samples, i.e. all samples are de-identified. In addition, TESR often request that analysis is blinded, such that IDs do not provide experimental details that could be a source of bias. Aggregated unidentified data are occasionally used to demonstrate the resource proficiency or to inform the range of expected concentrations in a particular sample type.

Start date: 04-01-2021

End date: 03-31-2026

Last modified: 02-07-2024

Copyright information:

Trace Element (TE) Shared Resource

Data Type

Types and amount of scientific data expected to be generated in the project: ***Summarize the types and estimated amount of scientific data expected to be generated in the project.***

Describe data in general terms that address the type and amount/size of scientific data expected to be collected and used in the project (e.g., 256-channel EEG data and fMRI images from ~50 research participants). Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be)

This project will generate both solution mode (with or without speciation) and solid mode (imaging) trace element measurements obtained by inductively coupled plasma mass spectrometry (ICP-MS) and several front-end separation techniques, including liquid chromatography for speciation and laser ablation for elemental imaging.

We expect to collect data from many hundreds of research specimens from multiple models (urine, toenails, blood, hair, teeth, internal organs), generating hundreds of datasets, ranging from smaller (50-100 MB) datasets for solution mode and speciation, to larger multi-dimensional datasets (2-3 GB) for solid mode analysis.

The following data files will be produced in the course of the project:

(1) Solution mode trace element analysis data exported from the ICP-MS Agilent Mass Hunter software as .csv files (raw data, concentrations, quality control and calibration statistics) and combined, along with a sample manifiest, all sample preparation steps and ICP-MS calibration preparation information) into multi-sheet Microsoft Excel (.xlsx) files.

(2) Elemental imaging (solid mode) ICP-MS or ICP TOFMS data are exported from the Agilent MassHunter or Vitesse ICP TOFMS software. Raw data (.csv and .vit files) will be transformed by the Iolite 4 software package, where they are baseline corrected and raw counts per second are quantified into parts per million with measurements from certified reference materials. The processed datasets are used for statistical analysis. To protect research participant identities, de-identified sample identities are used.

Imaging data can be supplied to the client either as fully processed .io4 files (if the client has this software) or as (raw) .vit, .dat or matrix files for use with Microsoft Excel. Images are also routinely

provided as .jpg, .pdf and .tiff. Imaging data ranges from 200 MB to 2GB per file and users generate on average 10-20 files.

Scientific data that will be preserved and shared, and the rationale for doing so: ***Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.***

Based on ethical and legal considerations, data produced by solid mode analysis will be shared among the elemental imaging community hosted by the TE shared resource for method development purposes only. In all cases, data will be shared after publication and associated embargo periods by the owner.

Solution mode data is collected as separate worksheets in a Microsoft Excel workbook, and maintained on the TE servers for use by the client only. Data processing, correction for sample preparation and dilution steps and quality control checking are conducted, and results and Quality Control metrics are recorded as worksheets in the workbook. Solution mode data will not be shared by the TE shared resource, which falls under the data management practices of the parent funding that supports fee-per-service analysis.

Metadata, other relevant data, and associated documentation: Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

To allow interpretation and publication of the data, metadata, sample preparation protocols and associated workbooks will be created and shared with the TE shared resource users. Metadata include instrument settings and parameters, and protocols used in processing raw data. All stages of sample preparation and analysis are recorded electronically in Microsoft Excel spreadsheets. Analytical balances are interfaced to laptops to transfer data without the risk of transcription errors.

Related Tools, Software and/or Code

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

For solution mode data, no specialized tools or software are needed. Data is collected as .csv files and converted into separate worksheets in a Microsoft Excel workbook, and all analyses are

recorded electronically in spreadsheets.

For solid mode (imaging) data, data will be made available in .io4 format for, which requires the use of the Iolite software application to be accessed and manipulated. These tools are fee-based, proprietary software. Alternative access to data for shared resource users without access to this software, data will be exported as Microsoft Excel matrix files. Conversely, raw data can be accessed via open-source freeware available online (namely Sam's microanalysis toolkit).

Standards

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources, and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist

Data processing, correction for sample preparation and dilution steps and quality control checking are conducted, and results and Quality Control metrics are recorded.

Data Preservation, Access, and Associated Timelines

Repository where scientific data and metadata will be archived: Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived.

Data storage and backup are the responsibility of the client generating the data. The workbook and all accompanying files are saved into the TE Core directory on the Earth Sciences GEO database, which is backed up daily to the Dartmouth Central servers.

How scientific data will be findable and identifiable: Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

Question not answered.

When and how long the scientific data will be made available: Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long

data will be available.

Question not answered.

Access, Distribution, or Reuse Considerations

Factors affecting subsequent access, distribution, or reuse of scientific data: NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing.

All data is transfer back to the client who requesting the study once it it finished. The client who owns the data is the responsible where they choose to store their data. This includes storage duration, access to this data and long term cost of making this data available.

Whether access to scientific data will be controlled: State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).

Data that cannot be shared or limited in what can be shared such as stated in IRB requirements, patient identifier information or consent requirements will need to establish data sharing agreements.

Protections for privacy, rights, and confidentiality of human research participants: If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

TEA analyses are not considered as human subjects research because no personal identifiers or other individual-specific data is supplied with the samples. Aggregated unidentified data are occasionally used to demonstrate the resource proficiency or to inform the range of expected concentrations in a particular sample type.

Oversight of Data Management and Sharing

Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

Tools for establishing a DSP along with links to NIH information sites are provided and easily accessible to clients. The continual monitoring of publications of investigators will help ensure compliance is being meet .
