EXAMPLE FOR THREE DATA TYPES: CLINICAL, PROTEOMICS, METABOLOMICS

ELEMENT 1: DATA TYPE;

A. Types and amount of scientific data expected to be generated in the project:

The proposed research will include three data types: clinical data, proteomics data and metabolomics data. The final clinical dataset will include self-reported demographic data and the results of study-related tests for loiasis and other filarial infections. The proteomics data will include raw LCMS/MS reads from biological samples prepared from each of the 354 research subjects. The metabolomics data will include raw LC-MS and GC-MS files from subject samples. Tabular data of hits will also be saved in .csv formats.

B. Scientific data that will be preserved and shared, and the rationale for doing so:

The Omics data will be made publicly available stripped of demographic information, and the de-identified clinical data will be made available by request.

C. Metadata, other relevant data, and associated documentation:

Protocols and details regarding instrument settings, data transformation and analysis will be made available in the accompanying plain text README document.

ELEMENT 2: RELATED TOOLS, SOFTWARE, AND/OR CODE

Data collection will be done in REDCap and data analysis will be performed using Microsoft Excel and R. An institutional instance of Box will be used for active data storage, which is HIPAA-compliant. For the proteomics data, resulting spectra will be converted to Mascot generic format (MGF) files using Proteome Discoverer v2.1.0.81. For the metabolomics data, raw structural information about GC-MS features will be obtained through spectral matching with the NIST 14 spectral library. Mass Profiler Professional software (Agilent) for GC-MS data and Analyst (AB Sciex) for LC-MS/MS will be used to assign peaks to raw ion chromatograms. No specialized tools or software will be needed to access or reuse the shared datasets, which will be available via a public repository (de-identified omics data), or via email after the completion of a DSA as a .csv file (de-identified clinical data).

ELEMENT 3: STANDARDS

A data dictionary will be provided for the clinical dataset that defines column headers, units of measurement and other pertinent metadata as necessary to understand and reuse the dataset. For the proteomics datasets, we will follow the Minimum information about a proteomics experiment (MIAPE) standard which was created by the Proteomics Standards Initiative of the Human Proteome Organization. For the metabolomics datasets, we will follow the guidelines dictated by the Metabolomics Standards Initiative (MSI). Both of the repositories selected for the omics data streams comply with the accepted standards for their field.

ELEMENT 4: DATA PRESERVATION, ACCESS, DISTRIBUTION AND ASSOCIATED TIMELINES

A. Repository where scientific data and metadata will be archived:

Concomitant with publication of the results of the study, participant level data that have been stripped of demographic information will be published as supplementary data and/or made publicly available (with restricted access as laid out in section 5 below) in the PI's institutional data repository.

Raw proteomics data and accompanying metadata will be made publicly available through the ProteomeXchange data repository for at least 10 years. Raw metabolomics data and accompanying

Commented [JB1]: *These example DMS Plans are provided for educational purposes to assist applicants with developing Plans but are not intended to be used as templates and **their use does not guarantee approval by NIH. Do not copy/paste this Plan without modifying it to reflect the types of data that are expected to be generated through your project.***

Note that the example DMS Plans may reflect additional expectations established by NIH or specific NIH Institutes, Centers, or Offices that go beyond the DMS Policy. Applicants will need to ensure that their Plan reflects any additional, applicable expectations (including from NIH policies, ICO policies, or as stated in the FOA).

In addition, these examples may reflect resources or policies that are in place at other institutions but that are not necessarily available at the University of Iowa. If needed, investigators can contact Research Data Services (lib-data@uiowa.edu) if they have questions regarding how to best complete their DMS Plan.

Commented [BJY2]: *Sample plan created by the NIH DMSP Guidance for Data Support Services Working Group https://osf.io/euaty*

metadata will similarly be made publicly available (10+ years) from the Metabolomics Workbench data repository which is accessible by metabolite data aggregator, MetabolomeXchange.

Scripts and coding workflows will be made available via GitHub.

B. How scientific data will be findable and identifiable:

The PIs institutional data repository will mint a DOI for datasets. Omics data will be assigned a persistent unique identifier during repository deposition and this PID will be included in relevant publications. All datasets deposited in the institutional repository are registered with Datacite, and are indexed by multiple search engines. These data deposit records also include links to the associated publications.

C. When and how long the scientific data will be made available:

The institutional repository will continue providing access to data sets for at least 10 years, or as long as the repository exists. The data repository also commits to maintaining at least one copy in the cloud in case of data loss. The ProteomeXchange and Metabolomics Workbench data repositories will keep data publicly available for at least 10 years.

ELEMENT 5: ACCESS, DISTRIBUTION, OR REUSE CONSIDERATIONS

A. Factors affecting subsequent access, distribution, or reuse of scientific data:

Access to the data will be restricted due to personally-identifiable information and personal health information. Results, data, and collections will be made available to other researchers in a timely basis following protocols to remove personally-identifiable information. Sensitive and confidential data collected will be treated following HIPAA regulations and IRB guidelines, and an added layer of security will be implemented by separating identifiable data from clinical data.

B. Whether access to scientific data will be controlled

The omics data will be made publicly available to users through [insert name] repositories. The clinical data will be stripped of most personal identifiers but will contain some indirect identifiers (such as age, sex, and location) in order to maintain the scientific value of the results.

C. Protections for privacy, rights, and confidentiality of human research participants:

To protect subject privacy, access to the clinical data will be controlled by the PI and released to users following the completion of a Data Sharing Agreement (DSA). The DSA will include instructions to: (1) use the data only for research purposes; (2) not identify any individual participant; (3) keep the data secured at all times; and (4) destroy or return the data after analyses have been completed.

ELEMENT 6: OVERSIGHT OF DATA MANAGEMENT AND SHARING

The PI of the proposal will make the plan available to all personnel involved in the project. The PI will be responsible for ensuring faithful adherence to the DMS Plan and revising the plan annually, as the research project evolves.