EXAMPLE DATA MANAGEMENT AND SHARING PLAN (in compliance with SF-424 Forms H)

EXAMPLE FOR TECHNOLOGY DEVELOPMENT PROJECT (CLINICAL)

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 [JB2]: *Example from NICHD https://www.nichd.nih.gov/sites/default/files/inline-files/ExampleDMSPlan_Human_ClinicalGenomics_NIHFormatPage_v2.pdf*

ELEMENT 1: DATA TYPE

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

1000 participants will be enrolled in the study. Clinical data will be obtained from the site electronic health record. Clinical data include demographical data, insurance status, medical history, medications, lab tests performed by the clinical site and central laboratory, and physical exams data, among other data pertinent to the study. Clinical data will be collected at varying frequencies (daily for up to 2 weeks then monthly for up to 12 months following informed consent). Research laboratory data include results of cytokine analysis and other immunoassays from biospecimens.

Clinical and laboratory data will be captured by each site into the REDCap secure electronic data capture system (EDC). Each user will be given role specific access to the EDC, and access will be controlled by granting users individual usernames and passwords. Research-specific labs will be recorded into the Central lab's database and merged with the final individual level data for the study prior to submitting to NICHD DASH. Clinical data from case report forms will be preserved and shared. Clinical data will be collected at varying frequencies (daily for up to 2 weeks then monthly for up to 12 months following informed consent).

Genomic data will be individual level data. Analysis of genomic DNA will link genotype to phenotypic information obtained as part of this study. Genomic data will include Whole Exome Sequencing (WES), shallow Whole Genome Sequencing (WGS), DNA methylation arrays (EPIC) for a total of 10 TB data. Genomic data will be obtained from a single blood draw taken on the first study visit following informed consent.

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

Clinical data that will be preserved and shared are demographical data, insurance status, medical history, medications, lab tests performed by the clinical site and central laboratory, and physical exams data, among other data pertinent to the study. Clinical data from case report forms will be preserved and shared. Research laboratory data that will be preserved and shared include results of cytokine analysis and other immunoassays from biospecimens. Genomic data that will be preserved and shared will include WES, shallow WGS, and DNA methylation arrays (EPIC). Clinical data sets will be submitted to DASH in .CSV format. Genomic data sets will be provided using FASTQ, CRAM, and VCF formats.

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

The protocol, sample informed consent, case report forms, data dictionary, and code book will be made accessible in data repositories where data are shared. For data submitted to DASH, variable-level metadata will be provided using the DASH Codebook, which is a templated data dictionary, and will include details of Common Data Elements, definitions, and standards used for data collection and sharing.

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

Clinical and laboratory data will be collected in the electronic data capture system (REDCap) and analyzed using open-source statistical packages in R. For genomic data analysis, containerized open-source workflows that are functionally equivalent to GATK pipelines will be used to call variants in this project and are or will be available through NICHD's GitHub repository.

ELEMENT 3: STANDARDS:

Data will be standardized to CDISC format whenever possible. Medical laboratory data will be standardized using LOINC. Shared data will be deidentified, and original data will be maintained at the investigator's institution. This research project will use the Pediatric and Pregnancy-related COVID Common Data Elements available via the Disaster Research Response (DR2) Resources Portal and the NIH Common Data Elements Repository, as required by the FOA. Genomic data will be in CRAM format with alignment to latest reference genome and variant files in VCF format.

ELEMENT 4: DATA PRESERVATION, ACCESS, AND ASSOCIATED TIMELINES

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

DASH (Clinical and laboratory data); dbGaP/Sequence Read Archive (SRA, Genomic data)

B. How scientific data will be findable and identifiable:

Clinical and laboratory data will be findable and identifiable using a DOI created by DASH. Genomic data will be findable and identifiable using a dbGaP study accession number and SRA sequence record accession numbers generated by NCBI.

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

The study team will submit an initial batch of processed and cleaned clinical and laboratory data to DASH after 500 participants have been enrolled and any remaining clinical data within 2.5 years of the last patient's last study visit, 4 months prior to any publication date, or 6 months before the end of the performance period, whichever is sooner. The repository will make the data accessible within 3 years of last patient's last study visit or at the end of the performance period, or no later than the date of publication, whichever is sooner.

The study will be registered in dbGaP at the time of data cleaning and quality control measure initiation. Genomic data will be shared according to GDS policy, which requires that genomic data must be submitted within 3 months following data generation and released within 6 months of data submission to the repository or at acceptance of the publication, whichever is first. The first submission to dbGaP/SRA will be timed to the data generation for the first 500 samples, and the second submission will be timed to the data generation for the second 500 samples. Data will be preserved within the repositories for at least three years following the completion of the grant, as required by federal retention guidelines.

ELEMENT 5: ACCESS, DISTRIBUTION, OR REUSE CONSIDERATIONS

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

Clinical data will be shared with controlled access in DASH for general research use, as allowed by the participant's informed consents and the Institutional Certification. Genomic data will be shared with controlled access in dbGaP/SRA for general research use, as allowed by the participant's informed consents and the Institutional Certification.

B. Whether access to scientific data will be controlled:

Data will be controlled access with the General Research Use Data Use Limitation, as allowed by the informed consent and the institutional certification.

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

Informed consent documents used for the proposed clinical trial will include explicit language informing the participant or legally authorized representative that residual biological specimens, including DNA, may be stored in a biorepository for other scientific investigations. The informed consents will contain

language permitting secondary use with broad data sharing under controlled access with general research use restrictions in DASH (clinical data) and dbGaP/SRA (genomic data). Patients will not be contacted or re-consented for future sharing or accessing data through repositories.
Privacy and confidentiality protections consistent with applicable federal, Tribal, state, and local laws, regulations, and policies will be followed. Data will be deidentified by removing all 18 HIPAA identifiers prior to sharing, and the study will have a Certificate of Confidentiality from NIH.

ELEMENT 6: OVERSIGHT OF DATA MANAGEMENT AND SHARING:

Data will be submitted by a project data manager from the PI's project team. The data manager will oversee data collection, analysis, storage, and sharing. Compliance with the plan will be monitored by the PI routinely. The PI will conduct monthly meetings with key study personnel to ensure the timeliness of data entry and will review data to ensure quality of data entry. The PI will ensure data are submitted and shared according to this DMSP.