

Data Management and Sharing Plan

Element 1: Data Type:

A. 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.

The study will generate clinical, laboratory, and imaging data from 360 participants (180 mother-infant pairs) enrolled across three study sites. Data collected includes demographic and medical history, physical exam findings, laboratory results (urinalysis, angiogenic factors, uric acid, glycemia), pregnancy prolongation, and neonatal outcomes (e.g., birth weight, Apgar scores, anthropometrics, NICU admission, Bayley neurodevelopmental assessments at 3, 6, and 12 months). Imaging data includes 2D Doppler ultrasound images (.jpeg), 3D fetal thigh volumetric images (.vol), and real-time grayscale cardiac clips (.dcm/DICOM). Tabular datasets (~1–5 MB per participant per timepoint) will be stored in REDCap and exported as .csv files, resulting in an estimated 10–20 GB total tabular data. Imaging files are estimated at ~30–50 GB across all formats.

B. 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.

All de-identified clinical, laboratory, and imaging data relevant to study endpoints will be preserved and shared, including tabular datasets of clinical/lab results, neurodevelopmental/Bayley scores, quantitative imaging measurements (e.g., pulsatility indices, volume measurements), and de-identified imaging files (jpeg, vol, dcm). Sharing these datasets enables reproducibility, meta-analyses, and secondary analyses by the scientific community, while maximizing the utility of the collected data in compliance with NIH and NICHD policies.

C. 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.

Metadata to be shared includes data dictionaries, variable definitions, coding manuals, data collection forms, imaging protocols, and data processing pipelines. Study protocol documents, informed consent templates (with language regarding broad data sharing), and descriptions of neurodevelopmental assessment instruments (e.g., Bayley Scales) will also be provided to facilitate interpretation and reuse.

Element 2: 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.

Most tabular data (.csv) can be accessed with standard spreadsheet or statistical analysis software (e.g., Microsoft Excel, R, SAS, SPSS). Imaging data will require standard viewers: .jpeg files (any image viewer), .vol files (e.g., TomTec, 3D Slicer—open source), and DICOM (.dcm) files (e.g., OsiriX, RadiAnt, 3D Slicer).

All required software is commercially available on open source. Links and instructions for accessing open-source tools will be provided with the data documentation.

Element 3: 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.

Clinical and laboratory data will be structured according to the Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM) where feasible. Imaging data will be annotated following DICOM standards. Metadata will follow the Data Documentation Initiative (DDI) standard. Definitions and variable coding will follow NIH Common Data Elements (CDEs) for pregnancy, neonatal, and neurodevelopmental research where available. In areas where consensus standards are not available (e.g., novel imaging measurements), detailed documentation and protocols will be provided to ensure usability.

Element 4: Data Preservation, Access, and Associated Timelines:

A. 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.

De-identified tabular data and documentation will be deposited in the NICHD Data and Specimen Hub (DASH), an NIH-supported, domain-specific repository for NICHD-funded studies. Imaging data will also be deposited in DASH or, if required due to file size or modality, in the NIH-supported National Institutes of Health Common Fund Data Ecosystem or the National Library of Medicine's National Center for Biomedical Imaging and Bioengineering (NLM/NBIB) repositories.

B. 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.

All datasets and metadata will be assigned persistent unique identifiers (DOIs) by the repository. Datasets will be indexed and discoverable via DASH and through NIH data discovery tools. Associated publications will reference these DOIs to ensure direct linkage.

C. 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 the time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

Data will be made available no later than the time of first publication of primary results, and in all cases by the end of the study performance period, in accordance with NIH policy. Data will remain available in the designated repositories for a minimum of 10 years after study completion, or longer per repository policy.

Element 5: Access, Distribution, or Reuse Considerations:

A. 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 participants will provide informed consent for broad data sharing, including future research use. Data will be de-identified in accordance with HIPAA and NIH guidelines. Potential limitations include the need to protect participant privacy and prevent re-identification, particularly for imaging and rare clinical presentations. Data use will be governed by repository data use agreements and policies.

B. 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).

Yes, access to de-identified individual-level data will be controlled via DASH or the selected repository. Investigators will be required to submit a data access request and agree to terms of use, including privacy protections and limitations on re-identification.

C. 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).

All data will be de-identified prior to sharing, with removal of direct identifiers and application of data suppression or generalization as needed to prevent re-identification. Certificates of Confidentiality will be obtained. Data sharing will comply with institutional IRB requirements and all relevant regulations. Only data for which broad sharing consent was obtained will be shared.

Element 6: 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).

Compliance with this Plan will be monitored by the study's Data Manager and Principal Investigator, with oversight from the institution's Office of Research Compliance and the Data and Safety Monitoring Board (DSMB). Regular reviews of data collection, de-identification, and sharing processes will be conducted at quarterly intervals. All data sharing activities will be documented, and compliance reports will be submitted annually to the funding agency as required.