

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.

This project will perform secondary data analysis on kidney magnetic resonance imaging (MRI) data to determine the parenchymal kidney volume. Analysis will be done on the subjects with available imaging data within the database of Genotypes and Phenotypes (dbGaP). The following dataset will be produced during this project: Clinical dataset of the estimated kidney volume.

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.

This secondary data analysis is not generating new data measurements. The data are already available through the dbGaP. The kidney volume dataset produced from the methods developed during this study will be made available for further research.

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.

Kidney volume will be estimated by a neural network trained using axial slices of dbGaP full-body MRI data. Manually created training data of known kidney volume from 95 subjects will be used to train and validate the model. Design choices and hyperparameter selections will be documented. The processing pipeline and related quality control metrics and thresholds will be defined and documented in a publication quality methods section.

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.

MRI images will be accessed using the Horos DICOM Viewer software available on Macintosh computers. The neural network will be implemented using Python. Estimated kidney volumes will be output into a tab-separated file that can be viewed by most text editing software.

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.

The data that will be used for the proposed secondary data analysis is already available on the dbGaP. The documentation for the generation of the data is also available. The estimated kidney length and volumes will be reported in the community standard of centimeters and milliliters.

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.

Estimated kidney volumes will be made available within the dbGaP as a sub-study. Identified study metadata will be supplied in the associated publication supplement as well as a supporting document in the dbGaP sub-study.

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.

The data will be associated with the subjects within the dbGaP. Links to the parent and sub-study datasets will be included in the publication.

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.

The data will be made available through the dbGaP following the standard review and release protocols. The data will be made available as soon as possible. Scientific data included in published manuscripts will be made available at the time of publication; all other scientific data will be made available no later than the end of the award.

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.

Access to the data will require investigators to register for a dbGaP account and submit an application providing a summary of the research you intend to conduct, the dataset you require, and a description of any new data or variables your research will generate. Data will then be made available following application approval.

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).

Data access will be controlled through the dbGaP. This will allow for association of kidney volume estimates produced from this study with the other clinical, imaging, and genomic data for the associated research participant.

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).

This study proposes to use only data that has already been de-identified. Shared data generated from this project will report subject- specific kidney volumes by using the study assigned ID, which has already been de-identified.

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).

The Principal Investigator for this project, Dr. ABC, will ensure that this Data Management and Sharing (DMS) Plan is followed. The institutional official (title and role), will be responsible for oversight of compliance with the accepted DMS Plan. Compliance will be evaluated annually during the award period and progress towards the plan's DMS activities will be included in the annual Research Performance Progress Report (RPPR) submitted to the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Project Officer. At the project conclusion, the final progress report will summarize how the DMS objectives were fulfilled and provide links to the shared dataset(s).