MWCCS Data Management and Sharing Plan for Ancillary Grants

Version 1.0

Effective January 19, 2022

PREAMBLE FOR INVESTIGATORS

This document addresses expectations and requirements for data sharing and establishing a written plan related to a proposal submitted to the MWCCS by investigators internal or external to the MWCCS, including proposals to be nested within the MWCCS consortium as an ancillary project. The proposal might (1) solely use existing MWCCS data, or (2) use existing data and are also collecting/generating additional new data within the MWCCS (e.g., new questionnaires, genetic studies using existing samples, imaging studies etc.)

This document was generated using the respective policies for NIH data management and sharing (DMS), and Genomic Data Sharing (GDS), as guidance for implementation. The draft language in the suggested template below can be used by proposing investigators when developing and including the required DMS plan (and GDS, where applicable) as part of a grant submission to NIH. It is anticipated that the draft language will be modified to appropriately reflect the activities proposed in a grant application. For example, grant applications proposing new 'omics' data, or generation of new imaging data, should plan to revise appropriate sections describing such data and the tools/software needed to analyze them.

In accordance with the 2023 NIH Data Sharing and Management Policy (NOT-OD-21-013, effective date is January 25, 2023), researchers are required to submit a Data Management and Sharing Plan outlining how scientific data and any accompanying metadata will be managed and shared, and comply with the Data Management and Sharing Plan approved by the funding Institute or Center. The DMS Policy applies to all research, funded or conducted in whole or in part by NIH, that results in the generation of scientific data. In this policy, NIH expects that all scientific data used for the funded proposal is to be made available at the time of publication or at the end of the funded period, whichever comes first.

In accordance with NIH DMS and GDS policies, which accommodate plans that maximize access for sharing scientific data, the MWCCS consortium has taken into account *the sensitive data collected and the vulnerabilities of those studied living with HIV, and local consent requirements restricting the release of data by many participants that increase the risk of deductive disclosure that limit full and open sharing*. As a result, MWCCS utilizes a plan for controlled release data management and sharing. In summary, all data (including any new data collected–or generated– as part of an ancillary grant) must be shared through the consortium-established concept sheet submission process. Concept sheets submitted to the MWCCS will include appropriate details as to how proposing investigators will be responsible for either: 1) depositing their analytic dataset directly to the Data Analysis and Coordination Center (DACC) for controlled release (this dataset will need to be given to the DACC at or before the time of manuscript submission to the Executive Committee), or 2) providing secure, long-term storage for their analytic dataset and submitting a data access plan to the DACC that details how data will be shared for approved research concepts.

NOTES/REFERENCES:
Investigators should review the references below to familiarize themselves with the components of the DMS plan.

NIH Guidance and FAQs

- DMS Website: https://sharing.nih.gov/data-management-and-sharing-policy
- FAQs: https://sharing.nih.gov/faqs#/data-management-and-sharing-policy.htm
- Common Data Elements: https://cde.nlm.nih.gov/home
- Genomic Data Sharing Policy: https://sharing.nih.gov/genomic-data-sharing-policy/about-genomic-data-sharing

Example DMS Plans

- https://datasharing.iupui.edu/nih-dms-plan-guidance.html
- https://www.nimh.nih.gov/funding/managing-your-grant/nimh-data-sharing-for-applicants-and-awardees#4

ATTACHMENT (MWCCS suggested template and language for a DMS Plan)

DATA MANAGEMENT AND SHARING PLAN (2 pages max)

Element 1: Data Type

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

TO BE COMPLETED BY INVESTIGATOR: Use the draft language below to describe the existing data provided by the MWCCS that will be used in your proposal.

Analysis of Existing Data

The proposed research will include secondary analysis of existing MWCCS data. The MWCCS is a multi-center, prospective, observational study that currently includes thirteen clinical research sites (CRS) and a compilation of data from over 10,000 former and current participants. Participants at each site are asked to take part in research study visits that include semi-structured interviews, clinical examinations, and specimen collection. All data are maintained in relation to a participant identifier (i.e., CCSID). Interview data are directly entered into a study-specific database managed by the DACC. Clinical exam and laboratory data are collected on paper forms, and entered by study staff into the study-specific database.

Data included in the analytic dataset provided for the proposed study include: (*INVESTIGATORS: modify as needed based on your proposal*)

Interview: Data collected in the interview include: socio-demographic information; health care utilization; general medical histories and prescription medication use including antiretroviral medications; use of alcohol, tobacco and drug use; sexual risk behaviors; beliefs regarding HIV and its treatment; and psychological status.

Clinical Outcome Confirmation/Ascertainment: Key self-reported clinical outcomes are confirmed by medical record abstraction (e.g., hospitalizations, cardiovascular disease (CVD) events, and malignancies).

Clinical Examinations: Brief physical examinations include standardized assessment of vital signs (e.g., blood pressure); anthropometric measures; frailty assessments; and a gynecological examination that includes cervical cytology (women only). Participants may undergo assessment for hepatic steatosis and fibrosis FibroScan® measures; pulmonary function testing; and transthoracic echocardiogram.

Laboratory Assessments and Specimen Collection: Participants are asked to provide whole blood, urine, rectal and pharyngeal, saliva, and stool samples, as well as a small sample of hair from the scalp for repository storage and laboratory testing, female participants are also asked to provide cervico-vaginal swabs and lavage fluid.

Neurocognitive Assessments: All participants are asked to complete a full neurocognitive battery every two-years.

Contextual Data: All MWCCS participants are asked to provide their residential address for geocoding and are assigned a census block group by site staff using ArcGIS™. A limited data set that contains only the participants' census block group number (FIPS) and the CCSID is securely transferred to the UNC CRS where it is linked to census-linked data sets, such as the American Community Survey, Decennial Census, American Housing Survey, and Annual Economic Survey, to create group-level variables that describe the locations where participants live. The UNC CRS creates individualized contextual datasets that only contain the CCSID, and the value of the specific group-level variables requested by the investigator (e.g., percent of individuals under the federal poverty line).

Genomic Data: Data from the prior WIHS GWAS analysis are currently housed at NYU under the stewardship of Dr. Aouizerat (through the San Francisco CRS site). Data from the prior MACS GWAS are currently housed at the DACC under the stewardship of Dr. Duggal. These data have been deposited with dbgap. We will comply with the NIH genomic data sharing policy. Submitted data will conform with relevant data and terminology standards. All other genomic data including RNA sequencing data and DNA methylation array and sequencing data will be shared as anonymized individual-data to the NIH funded Gene Expression Omnibus (GEO) repository (http://www.ncbi.nlm.nih.gov/geo/index.cgi), consistent with applicable laws and regulations

Collection/Generation of New Data
TO BE COMPLETED BY INVESTIGATOR: Describe the type of de novo collection or data generation. This may include new clinical assessments, surveys or data that will be generated using existing data (e.g., biomarker testing of existing specimens).

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.

Analysis of Existing Data
Existing MWCCS data include sensitive data that increase the risk of deductive disclosure of participants, as well as data from individuals who are unable to provide consent for open access data sharing. Scientific data generated through the proposed research will be preserved and made widely available through the MWCCS controlled access process by submission of a research concept sheet.

Collection/Generation of New Data
Prospective data to be generated include sensitive data that increase the risk of deductive disclosure that prevent open access data sharing. Scientific data generated through the proposed research will be preserved and made widely available through the MWCCS controlled access process by submission of a research concept sheet.

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.

Analysis of Existing Data
1. Copies of study data collection instruments (https://statepi.jhsph.edu/mwccs/data-collection-forms/) and manual of operations (https://statepi.jhsph.edu/mwccs/manual-of-operations/) are available on the MWCCS public website.
2. Metadata/overview of controlled access variables are available in MWCCS summary files, which are available upon approved Concept Sheet release under the MWCCS controlled access policy.

Collection/Generation of New Data
Describe metadata access for de-novo collection.

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.

Analysis of Existing Data
For interview, laboratory and clinical exam data:
Datasets can be requested - through the CCS controlled access process– in the format desired by the investigator including ASCII text file or CSV file. Syntax used and outputs from software or programs employed to analyze data are to be maintained for disclosure on request.

For other types of data
Describe any additional data elements that would require additional tools or software (e.g. imaging, genomics). Syntax used and outputs from software or programs employed for the derivation of novel variables are to be maintained for disclosure on request.

Collection/Generation of New Data
Describe any additional data elements that would require additional tools or software. Syntax used and outputs from software or programs employed for the derivation of novel variables are to be maintained for disclosure on request.

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.

Analysis of Existing Data
The MWCCS is an observational study that collects data that address a wide-range of domains. While no common standard exists for all of the MWCCS protocols, numerous domains have common data standards that have been adopted, including:

- Collection of standardized psychosocial scales [e.g., Center for Epidemiologic Studies-Depression Scale (CES-D), Perceived Stress Scale (PSS-4)];
- Standardized reporting of laboratory testing results (e.g., eGFR) and clinical assessments (e.g., pulmonary function index of lung function type).

Elements which have common standards are provided in summary files, which combine and harmonize longitudinal data across forms, in order to provide key variables in the permissible format. These summary files increase ability to access and process data from multiple sources without losing meaning and increase accessibility of investigators to use the data as coded by experts most familiar with the data.

Collection/Generation of New Data
Describe new elements that will be collected/generated with commons data standards:
https://cde.nlm.nih.gov/home

- e.g., DICOM standards will be used for medical images. https://www.dicomstandard.org/current

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; see Selecting a Data Repository).

All studies
Data generated by the proposed study will be made available through the MWCCS controlled access process (https://statepi.jhsph.edu/mwccs/work-with-us). This approval includes receipt of summary files.

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 Studies
The scientific data will be available for request through the MWCCS website (https://statepi.jhsph.edu/mwccs/work-with-us).

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

Analysis of Existing Data
The MWCCS data that will be used in the proposed research already exists and is made available through the controlled access process outlined above. The research community will have access to the analytic dataset generated as part of this study at the time of manuscript publication. The data will be made available via the MWCCS controlled access process and will be available indefinitely, upon approval of a MWCCS Concept Sheet.

Collection/Generation of New Data
The MWCCS data that will be used in the proposed research already exists and is made available through the controlled access process outlined above. The research community will have access to the analytic dataset, including new data elements that were collected, or generated, for the proposed study, at the time of manuscript publication. The data will be made available via the MWCCS controlled access process and will be available indefinitely, upon approval of a MWCCS Concept Sheet.

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. See Frequently Asked Questions for examples of justifiable reasons for limiting sharing of data.

Analysis of Existing Data
Data will be made widely available through the MWCCS controlled access policy. Applicable factors include:

- Data includes numerous sensitive and confidential data elements which, when combined, could result in deductive disclosure;
- Existing consent limits the extent of data sharing for future-use to individuals who have undergone the controlled-access process; and
- Re-consent of prior participants – who are no longer active in the study– would be impossible.

Collection/Generation of New Data
Data will be made widely available through the MWCCS controlled access policy. Applicable factors include:

- Data includes numerous sensitive and confidential data elements which, when combined, could result in deductive disclosure;
- Existing consent limits the extent of data sharing for future-use to individuals who have undergone the controlled-access process; and

- Re-consent of prior participants – who are no longer active in the study– would be impossible
- Describe any re-use considerations for data that will be newly collected/generated (e.g., DNA methylation data can be re-used to analyze in relation to different outcomes).

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).

All Studies

As outlined above, data will be made widely available through the MWCCS controlled access policy.

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 Studies

All MWCCS datasets are created ensuring that no linkages can be made with participants and that no identifying health information is included. The MWCCS datacenter does not have any link between the Combined Cohort Study ID (CCSID) and personal identifying information kept at the local sites only.

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).

Please work with your institution to determine their recommended strategy for monitoring compliance and oversight.

Example 1:

The Office of Sponsored Programs at University X which will be administering this award has created a data management and sharing plan compliance system as part of their process for submitting the annual NIH progress report.

Example 2:

PI: data analysis, data reporting, data management training for study personnel
Research Technicians: data collection, data entry, record-keeping, metadata creation
Post-doc: data screening & processing, data analysis, prepare datasets for sharing, data deposit and/or dissemination, data reporting
Research Scientist: data analysis, data visualization, data management training for study personnel
Research Coordinator: record-keeping, metadata creation, prepare datasets for sharing