Data Management and Sharing Plan Update

Chad Murchison, PhD – The University of Alabama at Birmingham

5/1/2023 – NACC Spring Meeting

A Support Document for an ADRC's DMSP

- Scoped to address the six elements required by NIH
- Specifically designed for NACC data streams
- Developed for data updates e.g. new data types, UDS4
- <u>FOR USE ALONG WITH AN ADRC'S OWN DMSP</u>

Elements of a DMS Plan

- **Data type**
  - Identifying data to be preserved and shared
- **Related tools, software, code**
  - Tools and software needed to access and manipulate data
- **Standards**
  - Standards to be applied to scientific data and metadata
- **Data preservation, access, timelines**
  - Repository to be used, persistent unique identifier, and when/ how long data will be available
- **Access, distribution, reuse considerations**
  - Description of factors for data access, distribution, or reuse
- **Oversight of data management and sharing**
  - Plan compliance will be monitored/ managed and by whom

DMSP Development Timeline

- ✓Jan 12 – RFA-AG-24-001 announced
- ✓Jan 25 – All NIH applications must include 2-page DMSP
- ✓Feb 1 – Requirements EDC sub-group forwards draft to NACC for updates and approval
- ✓March 15 – NACC forwards DMSP for review by NIA PO's
- ✓April 15 – Current draft returned to NACC for posting
- Mid-May – Registration of DataCite DOIs by NACC for finalized NACC-centric DMSP
- June 14 – P30 applications due

Current Draft of the NACC DMSP

Guidance for Management and Sharing of NACC Data

Background:

This document is a guide for developing the Data Management and Sharing (DMS) Plan language specific to NACC and Uniform Data Set (UDS) data for the six recommended elements of an NIH DMS Plan. All ADRCs are required to submit a DMS Plan under the Final NIH Policy for Data Management and Sharing (NOT-OD-21-013). Each ADRC is **also expected** to include details for management and sharing of non-NACC and non-UDS data in their DMS Plan. Sample DMS Plans are available on the NIH site for Writing a Data Management & Sharing Plan.

Element 1: Data types

NACC data encompasses five primary types as of January 2023:

- **NACC Uniform Data Set (UDS)** of longitudinal standardized clinical and neurocognitive phenotypic data collected annually from participants at each ADRC.
- **Neuropathology** autopsy data collected at death; non-standardized.
- **Neuroimaging data** (MRI, amyloid and tau PET)
- **Biological Sample** data for fluid biomarkers.
- **Genotypic and Genomic data.**

Prospective subject counts are determined by individual Centers. Current subject counts are dynamically updated after quarterly freezes and are available by data type via the NACC website "data summary tables" under "Requesting data".

Data collection for the NACC UDS is ongoing from the associated ADRCs. Guidelines and administration tools for the collection of NACC UDS clinical data are determined by NACC through its associated task forces and working groups and are readily available for reference online via the "Forms & documentation" section of the NACC website under "Data Collection".

Element 2: Related Tools, Software and/or Code

Validation and quality control of UDS data occur during upload to the main NACC database maintained by the University of Washington.

NIA-funded ADRCs upload data through password protected membership portals. Consistency and quality of UDS data is evaluated using proprietary SAS scripts developed and maintained by NACC staff. Data that does not pass validation must be corrected by ADRCs to be included in the dataset.

Bulk submission data files can be reviewed by Centers for quality control outside the NACC framework using the open source NACCulator tool made available by the CTS-IT of the University of Florida on their GitHub site.

Element 3: Standards

Data dictionaries and metadata level standardization of the described NACC data types are freely available and outlined in the "Forms & documentation" section of the NACC website. Conventions around variable naming, data quality, and continuity are provided by NACC and described in these data dictionaries for all data types described in Element 1.

MRI and PET data for ADRC participants will be collected in accordance with the acquisition protocols outlined by Standardized Centralized Alzheimer's and Related Dementias Neuroimaging (SCAN). Non-SCAN compliant MRI and PET images and biomarker data are available and data dictionaries can be found online under the "Forms & documentation" in the "Data collection" section of the NACC website.

Training materials related to NACC UDS administration are also made available online. "Coding Guidebooks" and the "Instructions for the Neuropsych Battery" contain NACC guidelines for instrument administration that may differ from guidelines in instrument specific documentation. Examples include the MoCA, classic CDR interview administrations, NPI-Q administration and the FTLD module additions to the CDR.

Element 4: Data Preservation, Access, and Associated Timelines

ADRCs will utilize NACC as the data repository for UDS data collected from ADRC participants. Data housed in the NACC system includes ADRC site identifiers and participant identifiers unique for each participant. UDS data and metadata for all previously described data types are preserved through routine uploads of validated data to the NACC database systems.

Preservation of other data types are conducted by NCRAD (biological samples), LONI (raw neuroimaging data), and NIAGADS (genotypic data) with access for all data facilitated by NACC. Data types are aligned by mapping of site-specific participant IDs to NACC IDs using the "PTID to NACCID Map" in the ADRC Portal.

Other data requesters accept the terms of the NACC Data Use Agreement and can submit requests for data for formal review at the NACC website. Date request policies and expectations are outlined in the NACC Handbook and available via the "Data request process" section of the NACC website. Publications utilizing NACC data are required to acknowledge NACC grants and submit abstracts and manuscripts to NACC for review during the publication process as detailed in the "Checklist for authors" section of the NACC website.

ADRCs upload NACC data at regular intervals and its associated collaborators according to data types and metadata listed above. NACC performs quality control and derivation of variables requiring calculations on new data added to the repository on a quarterly basis referred to as a data freeze. Each data freeze is given a unique identifier to allow easier comparisons of data analysis.

Element 5: Access, Distribution, or Reuse Considerations

ADRC site level consent development is supported through NACC resources regarding consent best practices, which include language around data storage, access, and capacity assessment. "Best practices" documents can be found under the "ADRC resources" section of the NACC website. Other data requesters accept the terms of the NACC Data Use Agreement (DUA) and can submit requests for data for formal review at the NACC website. Date request policies and expectations are outlined in the NACC Handbook. Publications utilizing data uploaded to NACC are required to acknowledge NACC grants and submit abstracts and manuscripts to NACC for review during the publication process as detailed in the Checklist for Authors.

Access to scientific data is made available through the NACC data request process and available after the signing of a DUA. Storing and sharing procedures through NACC systems are designed according to NIH GCDMP practices and standards which are outlined on the NACC website and the terms of the Data Use Agreement.

All data users are required to complete and sign a NACC DUA prior to dataset access. The NACC Data Use Agreements are designed to ensure appropriate use of the data solely by the individuals identified in the data request, ensure NACC is informed of intended manuscript submission, and ensure that NACC (and all ADRCs submitting data to NACC) are appropriately acknowledged in any publications. All data users are required to complete and sign a Data Use Agreement prior to dataset access. All collaborators on a project requesting NACC data are bound by the terms of the DUA and can freely share provided datasets amongst each other. NACC data utilization is limited to the data streams outlined above.

NACC distributes deidentified datasets to protect the privacy of study participants. All Data Requests are reviewed by the NACC Data Use Committee prior to distribution. The NACC Data Use Committee determines if researchers are provided with the Investigator or Commercial data file. This review ensures that participant's consent to participate in commercial-based research activities is respected.

Element 6: Oversight of Data Management and Sharing

ADRCs can monitor their NACC and UDS performance, data quality, and data summarizations through monthly reports from NACC, which are also sent to NIA. ADRC member sites have specific membership credentials allowing access to their working datasets for uploading of site NACC data and confirming accuracy.

How to Use in Your Center's DMSP

The following Data Management and Sharing Plan is for non-NACC data collected by the UAB ADRC. For details on data sharing and stewardship of NACC and NACC-affiliate data including the NACC Uniform Data Set, NACC neuropathology data, NACC neuroimaging data, fluid biological sample data, and genomic/genotypic data refer to the NACC-approved reference DMSP [1].

REFERENCES CITED

1. National Alzheimer's Coordinating Center. April 15, 2023. Guidance for Management and Sharing of NACC Data. DOI: 10.6069 / #########

How to Use in Your Center's DMSP

The following Data Management and Sharing Plan is for non-NACC data collected by the UAB ADRC. For details on data sharing and stewardship of NACC and NACC-affiliate data including the NACC Uniform Data Set, NACC neuropathology data, NACC neuroimaging data, fluid biological sample data, and genomic/genotypic data refer to the NACC-approved reference DMSP [1].

REFERENCES CITED

1. National Alzheimer's Coordinating Center. April 15, 2023. Guidance for Management and Sharing of NACC Data. DOI: 10.6069 / #########

UW DataCite Prefix

Unique DMSP suffix

Writing Your Center's Unique DMSP for NIH

- Focus on the data management and sharing specifically for your center and institution
- Emphasize unique data streams not described in the NACC DMSP e.g. digital biomarkers
- Include details on NACC data hosted within your ADRC
- Be sure to include a prospective sample size in Element 1
- Cite the NACC DMSP as you would any other DOI citation
- NO HYPERTEXT!

NO HYPERTEXT!

Guidance for Management and Sharing of NACC Data

Background:

This document is a guide for developing the Data Management and Sharing (DMS) Plan language specific to NACC and Uniform Data Set (UDS) data for the six recommended elements of an NIH DMS Plan. All ADRCs are required to submit a DMS Plan under the Final NIH Policy for Data Management and Sharing (NOT-OD-21-013). Each ADRC is **also expected** to include details for management and sharing of non-NACC and non-UDS data in their DMS Plan. Sample DMS Plans are available on the NIH site for Writing a Data Management & Sharing Plan.

Element 1: Data types

NACC data encompasses five primary types as of January 2023:

- **NACC Uniform Data Set (UDS)** of longitudinal standardized clinical and neurocognitive phenotypic data collected annually from participants at each ADRC.
- **Neuropathology** autopsy data collected at death; non-standardized.
- **Neuroimaging data** (MRI, amyloid and tau PET)
- **Biological Sample** data for fluid biomarkers.
- **Genotypic and Genomic data.**

Prospective subject counts are determined by individual Centers. Current subject counts are dynamically updated after quarterly freezes and are available by data type via the NACC website "data summary tables" under "Requesting data".

Data collection for the NACC UDS is ongoing from the associated ADRCs. Guidelines and administration tools for the collection of NACC UDS clinical data are determined by NACC through its associated task forces and working groups and are readily available for reference online via the "Forms & documentation" section of the NACC website under "Data Collection".

Element 2: Related Tools, Software and/or Code

Validation and quality control of UDS data occur during upload to the main NACC database maintained by the University of Washington.

NIA-funded ADRCs upload data through password protected membership portals. Consistency and quality of UDS data is evaluated using proprietary SAS scripts developed and maintained by NACC staff. Data that does not pass validation must be corrected by ADRCs to be included in the dataset.

Bulk submission data files can be reviewed by Centers for quality control outside the NACC framework using the open source NACCulator tool made available by the CTS-IT of the University of Florida on their GitHub site.

Element 3: Standards

Data dictionaries and metadata level standardization of the described NACC data types are freely available and outlined in the "Forms & documentation" section of the NACC website. Conventions around variable naming, data quality, and continuity are provided by NACC and described in these data dictionaries for all data types described in Element 1.

MRI and PET data for ADRC participants will be collected in accordance with the acquisition protocols outlined by Standardized Centralized Alzheimer's and Related Dementias Neuroimaging (SCAN). Non-SCAN compliant MRI and PET images and biomarker data are available and data dictionaries can be found online under the "Forms & documentation" in the "Data collection" section of the NACC website.

Training materials related to NACC UDS administration are also made available online. "Coding Guidebooks" and the "Instructions for the Neuropsych Battery" contain NACC guidelines for instrument administration that may differ from guidelines in instrument specific documentation. Examples include the MoCA, classic CDR interview administrations, NPI-Q administration and the FTLD module additions to the CDR.

Element 4: Data Preservation, Access, and Associated Timelines

ADRCs will utilize NACC as the data repository for UDS data collected from ADRC participants. Data housed in the NACC system includes ADRC site identifiers and participant identifiers unique for each participant. UDS data and metadata for all previously described data types are preserved through routine uploads of validated data to the NACC database systems.

Preservation of other data types are conducted by NCRAD (biological samples), LONI (raw neuroimaging data), and NIAGADS (genotypic data) with access for all data facilitated by NACC. Data types are aligned by mapping of site-specific participant IDs to NACC IDs using the "PTID to NACCID Map" in the ADRC Portal.

Other data requesters accept the terms of the NACC Data Use Agreement and can submit requests for data for formal review at the NACC website. Date request policies and expectations are outlined in the NACC Handbook and available via the "Data request process" section of the NACC website. Publications utilizing NACC data are required to acknowledge NACC grants and submit abstracts and manuscripts to NACC for review during the publication process as detailed in the "Checklist for authors" section of the NACC website.

ADRCs upload NACC data at regular intervals and its associated collaborators according to data types and metadata listed above. NACC performs quality control and derivation of variables requiring calculations on new data added to the repository on a quarterly basis referred to as a data freeze. Each data freeze is given a unique identifier to allow easier comparisons of data analysis.

Element 5: Access, Distribution, or Reuse Considerations

ADRC site level consent development is supported through NACC resources regarding consent best practices, which include language around data storage, access, and capacity assessment. "Best practices" documents can be found under the "ADRC resources" section of the NACC website. Other data requesters accept the terms of the NACC Data Use Agreement (DUA) and can submit requests for data for formal review at the NACC website. Date request policies and expectations are outlined in the NACC Handbook. Publications utilizing data uploaded to NACC are required to acknowledge NACC grants and submit abstracts and manuscripts to NACC for review during the publication process as detailed in the Checklist for Authors.

Access to scientific data is made available through the NACC data request process and available after the signing of a DUA. Storing and sharing procedures through NACC systems are designed according to NIH GCDMP practices and standards which are outlined on the NACC website and the terms of the Data Use Agreement.

All data users are required to complete and sign a NACC DUA prior to dataset access. The NACC Data Use Agreements are designed to ensure appropriate use of the data solely by the individuals identified in the data request, ensure NACC is informed of intended manuscript submission, and ensure that NACC (and all ADRCs submitting data to NACC) are appropriately acknowledged in any publications. All data users are required to complete and sign a Data Use Agreement prior to dataset access. All collaborators on a project requesting NACC data are bound by the terms of the DUA and can freely share provided datasets amongst each other. NACC data utilization is limited to the data streams outlined above.

NACC distributes deidentified datasets to protect the privacy of study participants. All Data Requests are reviewed by the NACC Data Use Committee prior to distribution. The NACC Data Use Committee determines if researchers are provided with the Investigator or Commercial data file. This review ensures that participant's consent to participate in commercial-based research activities is respected.

Element 6: Oversight of Data Management and Sharing

ADRCs can monitor their NACC and UDS performance, data quality, and data summarizations through monthly reports from NACC, which are also sent to NIA. ADRC member sites have specific membership credentials allowing access to their working datasets for uploading of site NACC data and confirming accuracy.

NO HYPERTEXT!

Hypertext, Hyperlinks, and URLs

- Hyperlinks and URLs are only allowed when specifically noted in funding opportunities and/or form field instructions. It is highly unusual for a funding opportunity to allow links in Specific Aims, Research Strategy, and other page-limited attachments.
- Hyperlinks and URLs may not be used to provide information necessary to application review. Applications must be self-contained and reflect the information available at time of review.
- Reviewers are not obligated to view linked sites and are cautioned that they should not directly access a website (unless the link to the site was specifically requested in application instructions) as it could compromise their anonymity.
- When allowed, you must hyperlink the actual URL text so it appears on the page rather than hiding the URL behind a specific word or phrase (hypertext).
- **Note:** *During initial implementation of the new Data Management and Sharing Policy, NIH will provide leniency and will not withdraw applications that include hypertext in the DMS Plan. Beginning with applications submitted for due dates on or after May 25, 2023, failure to follow the standard instructions regarding hypertext may require NIH to withdraw your application from consideration.*

https://grants.nih.gov/grants/how-to-apply-application-guide/format-and-write/format-attachments.htm

Data Management and Sharing Plan Format

DMS Plans are recommended to be two pages or less in length.

NIH has developed an optional DMS Plan format page that aligns with the recommended elements of a DMS Plan.

Important: Do not include hypertext (e.g., hyperlinks and URLs) in the DMS Plan attachment.

https://sharing.nih.gov/data-management-and-sharing-policy/planning-and-budgeting-for-data-management-and-sharing/writing-a-data-management-and-sharing-plan

DMSP Resources

- NIH Data and Sharing Policy Site
  - Writing guidelines and example DMSPs
  - Budgeting
- Your local Office of Sponsored Programs or Research
- Our wonderful NIA Program Officers
- Watch for updates from NACC and the EDC Workgroup