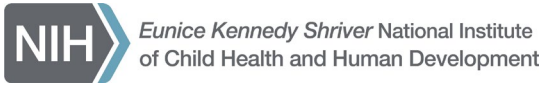
NICHD Data and Specimen Hub (DASH)

DASH Codebook Preparation and Submission

Webinar – June 22, 2022



Topics

1. NICHD DASH Overview
2. Data and Codebook Submission
    a. Requirements for Data Submission to DASH
    b. Codebook Preparation – Overview/Demo
        - DASH Codebook Template and User Guide
        - Populating the Template – Examples/Best Practices
    c. Data Preparation Tool – Demo
3. Resources for Data and Codebook Submission
4. Q&A

NICHD DASH Overview

DASH Data and Specimen Hub

- Centralized resource for researchers to share de-identified data from studies funded by NICHD
- Serves as a portal for requesting biospecimens from selected studies in DASH
- Launched in August 2015 and governed by the DASH Core Group
- Aims to accelerate scientific findings to ultimately improve human health

201 Studies

483 Data Requests

9 Studies Offering Biospecimens

50 Study Topics

70 Data Use Publications

11 Biospecimen Requests

https://dash.nichd.nih.gov

DASH Content

Top Study Topics

- Breastfeeding & Breast Milk *
- Cerebral Palsy
- Child Health *
- Early Learning
- Early Labor and Birth
- Fertility Problems
- High-Risk Pregnancy
- HIV/AIDS *
- Infant Care & Health *
- Infant Mortality
- Infertility & Fertility
- Labor & Delivery
- Newborn Screening
- Obesity & Overweight
- Pelvic Floor Disorders
- Pharmacology
- Pregnancy *
- Preterm Labor & Birth *
- Stillbirth
- Women's Health *

*biospecimens available

Currently Available Biospecimens

- Amniotic fluid
- Blood
- Breast Milk
- Buffy Coat
- Cord Blood (Buffy Coat, RBC, Plasma, Serum)
- DNA/RNA/Proteins
- Environmental Samples
- Erythrocytes (RBC)
- Hair
- Lymphocytes
- Meconium
- Nail
- Saliva
- Serum/Plasma
- Tissue samples
- Urine
- Vaginal Fluid

Last updated 06/20/2022

Benefits of Sharing Data in DASH

- Comply with NIH data sharing policies
  - Share data associated with a publication or share all study-related data in a timely manner, in accordance with any consent-based data use limitations and NIH data access review process
- Maximize reuse of data already collected through secondary analysis
  - Pool data with other study data to perform secondary analyses
  - 1.6 years is the average time to publication for findings from secondary analysis of DASH data
- Enhance the scientific value of NICHD studies through broad data sharing
  - One study in DASH (Consortium on Safe Labor) requested over 110 times in the past 6 years by investigators globally
  - Original study investigators will be acknowledged in new publications resulting from data reuse (requirement as per DASH policy)
- Jump start the careers of early-stage investigators by supporting hypothesis generation or secondary analysis of existing data

Data and Codebook Submission

Submission Requirements, DASH Codebook Preparation, Data Preparation Tool

DASH Data Submission Requirements* – Overview

1. **Institutional Certification for Sharing**
   Submitting institution must attest to *one* of the following:
   a. The study was conducted under an IRB determination of exemption or waiver of consent.
   b. The study was conducted as part of an IRB Approved Protocol; there is specific language describing sharing in the original consent AND/OR an IRB has reviewed the consent and determined that the data can be shared for research purposes
   c. The research was conducted with de-identified data that did not require consent

2. **Study Protocol and Instruments**
   Includes the background, rationale, objectives, design, methodology, statistical considerations, and organization of a clinical research study, as well as the instruments used for data collection

3. **Data De-identification Methodology**
   Summarizes the redaction, masking, and/or de-identification methods used to de-identify the data and biospecimen catalog as per HIPAA regulations (Safe Harbor method)

4. **DASH Codebook**
   Documentation of the study measures as well as the variables and their corresponding values (e.g., 0=No, 1=Yes)

5. **Biospecimen Catalog (if applicable)**
   Provides information about biospecimens collected from the study and stored in the NICHD Biorepository that are available for sharing through DASH

**See Appendix for more details*

DASH Data Submission – High Level Process Overview

Principal Investigator / DCC

DASH Curation Team

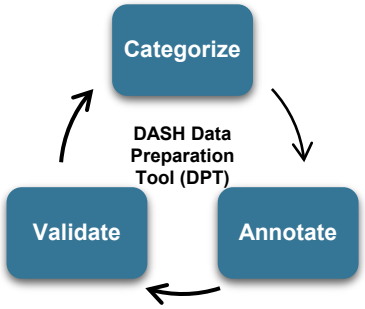
Part I: Prepare Study Documentation and Data

1. Gather data and required study documentation

2. De-identify study data and biospecimen catalog (if applicable, using the same participant ID code)

Part II: Submit in DASH

3. Prepare study data, DASH Codebook, study documents, biospecimen catalog (if applicable) for submission using the DASH Data Preparation Tool



4. Upload and submit study in DASH

- Institutional Certification
- De-identified study data
- Study Documents (study protocol, DASH Codebook, data collection instruments, de-identification methodology)

5. Submission complete, documented, and data de-identified properly?

No

Yes

6. Obtain NICHD DER Branch Chief/ DIR Director/ NCMRR Director Sign Off

7. Study Released in DASH

Part III: Review and Approve Study

DASH Codebook Preparation – Overview/Demo

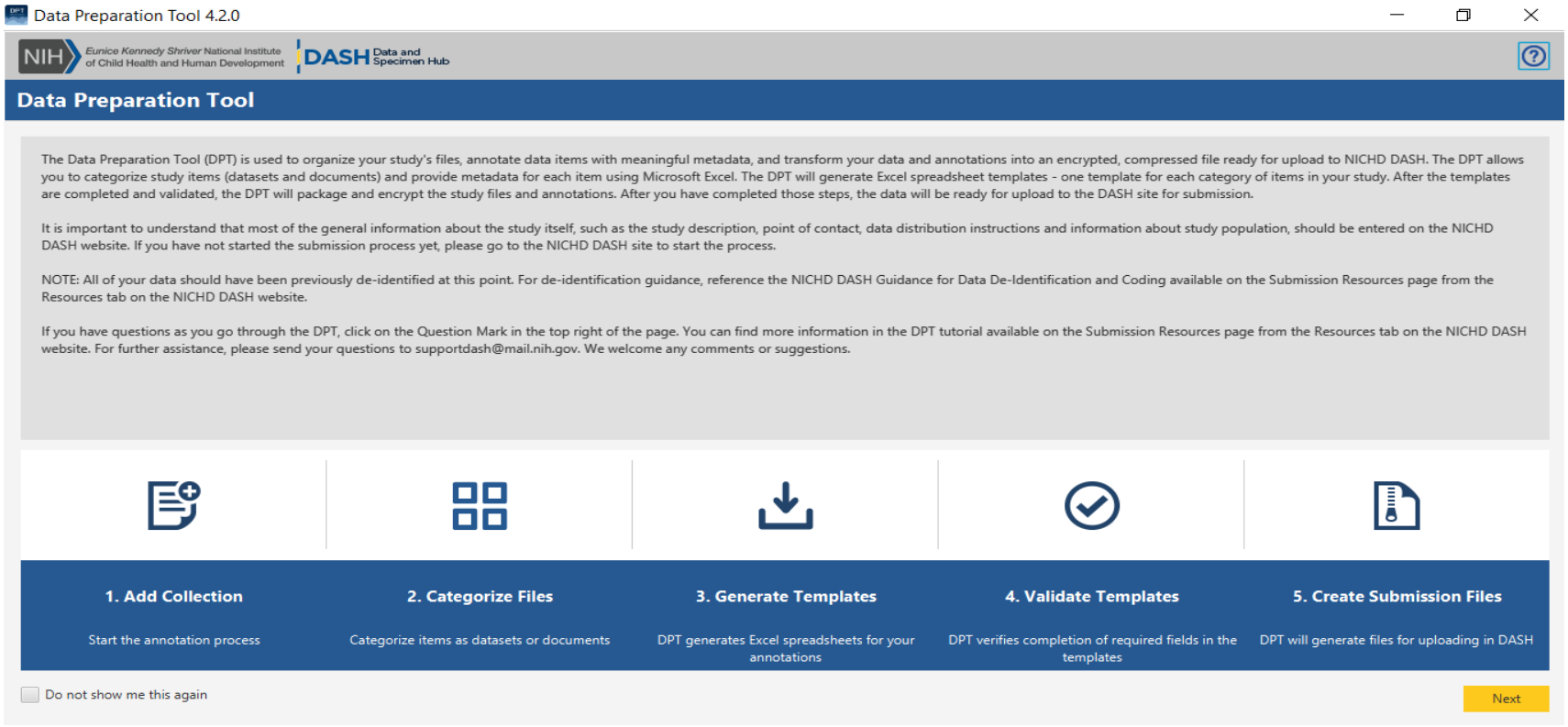
- **DASH Codebook Template:** Templated data dictionary that captures information about datasets, variables, and coded values for all data submitted to DASH for a given study. *Benefits of codebook include:*
  - Allows other researchers to understand and accurately interpret the variables and corresponding values in the study data
  - Standardizes metadata associated with variables submitted to DASH to enable cross-study comparison of datasets
  - Facilitates data submission, quality assurance, archiving, and sharing processes more efficiently in DASH
  - Provides recommended terminology standards for coding participants responses (*see Recommended Standards tab in the Template)*

- **DASH Codebook User Guide**: Guidance for submitters to populate the DASH Codebook template with variable information and includes best practices for completion of the DASH Codebook

- Both the Codebook Template and User Guide are available for download from the Submission Resources page in DASH

DASH Data Preparation Tool – Demo

- **DASH Data Preparation Tool (DPT):** Desktop tool for submitters to organize, annotate, and securely submit items (e.g., datasets, documents, and biospecimen catalog) associated with their studies
  - DPT is available for download from the Submission Resources page in DASH (Windows and Mac OS X)
  - Detailed instructions are included in the DPT Tutorial available for download



Resources for Data and Codebook Submission

DASH Resources

General DASH Resources

- NICHD DASH Policy: Outlines DASH policies and governance for data submission and request: https://dash.nichd.nih.gov/resource/policies
- DASH User Agreement: Outlines the terms and conditions users must agree to for using DASH: https://dash.nichd.nih.gov/resource/DASHUserAgreement
- Frequently Asked Questions: Provides answers to common user questions: https://dash.nichd.nih.gov/resource/FAQs
- DASH Tutorial: Step-by-step instructions on how to create a DASH account, search for and/or request data and biospecimens, and submit studies: https://dash.nichd.nih.gov/resource/tutorial
- Links to Educational Datasets (for training purposes): https://dash.nichd.nih.gov/resource/LinksToEducationalDatasets

Data and Codebook Submission Resources

- DASH Codebook Template
- DASH Codebook User Guide
- Study Submission Worksheet
- Institutional Certification Template for Data Submission
- Institutional Certification Template for Biospecimen Catalog Submission
- DASH Data and Biospecimen Catalog De-identification Guidance
- DASH Data Preparation Tool (Windows and Mac OS X)

DASH Team POCs

- DASH Program Lead: Dr. Rebecca Rosen: rebecca.rosen@nih.gov
- DASH Presenters
  - Susan Tenney, Lead Program Support
  - Alissa Fujimoto, Lead Data Manager
- Contact DASH Administrator at SupportDASH@mail.nih.gov for assistance or any questions about DASH

Appendix

1. Institutional Certification (Approval for Sharing)

Background

- A completed and signed Institutional Certification from the submitting institution is required before a study can be submitted to DASH
- The submitting institution must attest to *one* of the following:
  a. The study was conducted under an IRB determination of exemption or waiver of consent
  b. The study was conducted as part of an IRB Approved Protocol; there is specific language describing sharing in the original consent AND/OR an IRB has reviewed the consent and determined that the data can be shared for research purposes
  c. The research was conducted with de-identified data that did not require consent
- Obtaining approval for data and biospecimen sharing can delay submission, especially for completed studies; therefore, plan to obtain the Institutional Certification in advance

Recommendations

For Prospective Studies

- Ensure that language for broad data sharing is included in the informed consent documentation when planning the study
- Work closely with the IRB before the study closes to ensure that data sharing plans are consistent with informed consent procedures

For Retrospective (Completed) Studies

- Contact the DASH Program Lead to discuss sharing a completed study where the study IRB is no longer active

The Institutional Certification Templates for Data Submission and Biospecimen Catalog are available for download from the Submission Resources page in DASH: https://dash.nichd.nih.gov/resource/submission

2. Study Protocol and Instruments

Background

- Studies submitted to DASH must be accompanied by proper documentation to ensure meaningful use of the data for secondary analysis
- Study protocol and data collection instruments are required documentation for archiving in DASH
  - Study procedures are documented in the protocol and data collection instruments
    - Study protocol describes the background, rationale, objectives, design, methodology, statistical considerations, and organization of a clinical research study
    - Data collection instruments/case report forms are tools used to collect data in a clinical research study

Recommendations

For Prospective Studies

- When possible, use standard templates such as the NIH Protocol Templates for Clinical Trials

For Prospective and Retrospective (Completed) Studies

- Confirm that study protocols capture, at a minimum, essential information to ensure a basic level of understanding and usability for requesting investigators (e.g., study objectives, study design, participant recruitment/enrollment)
- Provide clear linkages between the instruments and resulting datasets/variables
- Ensure that proprietary instruments are not included in the submission to DASH

3. Data De-identification Methodology

Background

- Only submit de-identified study data or biospecimen catalog to DASH
- Investigators must strip all individually identifying information and assign random unique codes to the de-identified data, according to the following standards:
  - HHS Regulations for the Protection of Human Subjects ("Common Rule") and
  - Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule
- Refer to the DASH Data and Biospecimen Catalog De-identification Guidance before performing de-identification
- The de-identification methodology must be documented to summarize the redaction, masking, and/or de-identification methods used to de-identify the data

Recommendations

For Prospective Studies

- Maintain a log/record while the study is in progress of variables that will need to be de-identified
- Redact study data as appropriate by removing the identifier, preserving as much data as possible, and documenting steps taken to de-identify data

For Retrospective (Completed) Studies

- Consult with the DASH Curation Team, as needed, prior to de-identifying variables to ensure meaningful data are retained
- Document data de-identification procedures for traceability, replicability, and accountability

Contact the DASH Curators at supportDASH@mail.nih.gov with any questions

4. DASH Codebook

Background

- Investigators who plan to share data through DASH are expected to submit a DASH Codebook as part of their study submission
- The DASH Codebook is a data dictionary that captures information about datasets, variables, and coded values, for all data submitted for a given study
  - Essential for researchers to accurately understand, use, and analyze the variables and corresponding values in the data
  - Captures key variable information such as:
    - Labels/Definitions
    - Type (e.g., numeric, coded, or character)
    - Code-lists and associated values
    - Units of measurement

Recommendations

For Prospective Studies

- Maintain DASH Codebook throughout study data collection or maintain log/record of the variables and associated information needed to populate the DASH Codebook
- If using SAS software, consider using labels and formats to describe the data
- Refer to the Recommended_Standards tab for terminology standards that should be used to the extent possible when coding participant responses

For Retrospective (Completed) Studies

- Confirm that codebooks/data dictionaries contain, at a minimum, the required fields in the DASH Codebook template
- For study data in SAS format, any associated files (such as transport files) should accompany the datasets as well as any additional programming code that would facilitate the understanding and use of variables

The DASH Codebook Template and User Guide are available for download from the Submission Resources page in DASH: https://dash.nichd.nih.gov/resource/submission

5. Biospecimen Catalog (if applicable)

Background

- DASH facilitates investigator requests to access biospecimens from the NICHD Biorepository for select DASH studies and provides linkages between a study's data and corresponding biospecimens
- Biospecimen Catalog submitted to DASH must include the following essential fields:
  - Participant identifier
  - Biospecimen identifier
  - When the biospecimen was collected, the visit name, or visit ID
  - Type of biospecimen collected
  - Amount of biospecimen available
  - Age and sex of the participant
  - Any restrictions for use

Recommendations

For Prospective Studies

- Ensure that biospecimen sharing is explicitly stated in the informed consent
- Label the biospecimens and the study data with the same unique de-identified code and ensure biospecimen labels are free of any identifiers
- Conduct biospecimen checks to ensure the quality of stored samples

For Retrospective (Completed) Studies

- Perform multiple data quality checks between the biospecimen catalog and study data to confirm data accuracy/consistency, and ensure participant identifiers are correct and matched between sources

Guidance on data and biospecimen catalog de-identification is available for download at https://dash.nichd.nih.gov/download-api/resource?fileName=NICHD-DASH-Data-and-Biospecimen-Catalog-De-Identification-Guidance.pdf