Supporting Statement A for Clearance for

# The *Eunice Kennedy Shriver* National Institute of Child Health and Development

# **Data and Specimen Hub (DASH)**

Date: February 22, 2016

**Rohan Hazra, M.D., Chief, Maternal and Pediatric Infectious Disease Branch**

*Eunice Kennedy Shriver* **National Institute of Child Health and Human Development (NICHD)**

**National Institutes of Health**

**6100 Executive Blvd., Room 4B11**

**Bethesda, MD 20892-7510**

**Telephone: 301-435-6868**

**Fax: 301-496-8678**

**Email: hazrar@mail.nih.gov**

Check off which applies:
- X   New
- Revision
- Reinstatement with Change
- Reinstatement without Change
- Extension
- Emergency
- Existing

**Table of C**ontents

## A.    Justification

The NICHD Data and Specimen Hub (DASH) is being established by NICHD as a data sharing mechanism for biomedical research investigators. It will serve as a centralized resource for investigators to store and access de-identified data from studies funded by NICHD. Anyone can access NICHD DASH to browse and view descriptive information about the studies and data archived in NICHD DASH without creating an account. Users who wish to submit or request research study data must register for an account.

Information will be collected from those wishing to create an account, sufficient to identify them as unique Users. Those submitting or requesting data will be required to provide additional supporting information to ensure proper use and security of NICHD DASH data. The information collected from Users who register in NICHD DASH and submit and/or request data will be used to monitor data submissions and requests, to monitor Users' experience with DASH, and to notify interested recipients of updates to data stored in NICHD DASH.

The potential for public benefit to be achieved through sharing research study data for secondary analysis is significant. NICHD DASH supports NICHD's mission to ensure that every person is born healthy and wanted, that women suffer no harmful effects from reproductive processes, and that all children have the chance to achieve their full potential for healthy and productive lives, free from disease or disability, and to ensure the health, productivity, independence, and well-being of all people through optimal rehabilitation. Data sharing and reuse will promote testing of new hypotheses from data already collected, facilitate trans-disciplinary collaboration, accelerate scientific findings and enable NICHD to maximize the return on its investments in research.

## A.1    Circumstances Making the Collection of Information Necessary

This document contains information supporting a request for the Office of Management and Budget (OMB) to approve a clearance for the collection of information during registration, data submission, and data request procedures associated with the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) Data and Specimen Hub (DASH). Public Law 87-838 (enacted October 17, 1962) authorized the establishment of Institute of Child Health and Human Development for the conduct and support of research and training relating to maternal health, child health and human development, including research and training in the special health problems and requirements of mothers and children and in the basic sciences relating to the processes of human growth and development, including prenatal development. The information to be collected will be used for the purpose of identifying Users and ensuring proper use and security of NICHD DASH data.

NICHD conducts and funds over 2000 clinical research studies annually and a majority of the studies are conducted at various academic and research institutions across the U.S. as well as other countries. The data generated from these studies are under the purview of the study investigators and are not easily accessible due to challenges with data storage locations, formats and structure. To address these challenges and enable broader data sharing, NICHD established DASH (https://dash.nichd.nih.gov/) as a centralized resource for researchers to store and access de-identified data from studies funded by NICHD. It will allow NICHD funded extramural and intramural investigators to comply with NIH data sharing policies, by

enabling investigators to organize, store, and mine data from NICHD funded research studies for purposes of secondary research.

Establishing a central data sharing resource such as NICHD DASH also meets the objectives of various NIH and federal data sharing initiatives, including:

- *NIH Big Data to Knowledge (BD2K) Program (2012)* – Includes the Data and Informatics initiative aimed at facilitating the use of and maximizing the value of biomedical data by improving data sharing policies, cataloging research data
- *Federal Policy on Public Access (Feb 2013)* – The White House Office of Science Technology and Policy Memo on increasing access to the results of federally funded scientific research includes an objective to store digitally formatted scientific data enabling search, retrieve, and analyze capabilities
- *The White House Open Data Policy (May 2013)* – Mandates that 'data are released to the public in ways that make the data easy to find, accessible, and usable'

By facilitating data sharing, NICHD will promote the secondary use of research data already collected, reinforce open scientific inquiry, bring together investigators from multiple disciplines and ultimately advance the scientific mission of NICHD.

In order to enable data sharing through DASH, both User and study information will be collected from the Users of the system. User information stored in NICHD DASH is protected under the Privacy Act of 1974 (Pub.L. 93–579, 88 Stat. 1896, enacted December 31, 1974, 5 U.S.C. § 552a), which establishes a Code of Fair Information Practice that governs the collection, maintenance, use, and dissemination of personally identifiable information about individuals maintained in systems of records by federal agencies. Though the research data stored in NICHD DASH will be de-identified, risks to individuals, groups, or communities will be balanced carefully with potential benefits of the knowledge to be gained through NICHD DASH. To protect the privacy of research participants and the confidentiality of their data, the NICHD DASH Data Access Committee will review the request for proposed research study and monitor the use of NICHD DASH data.
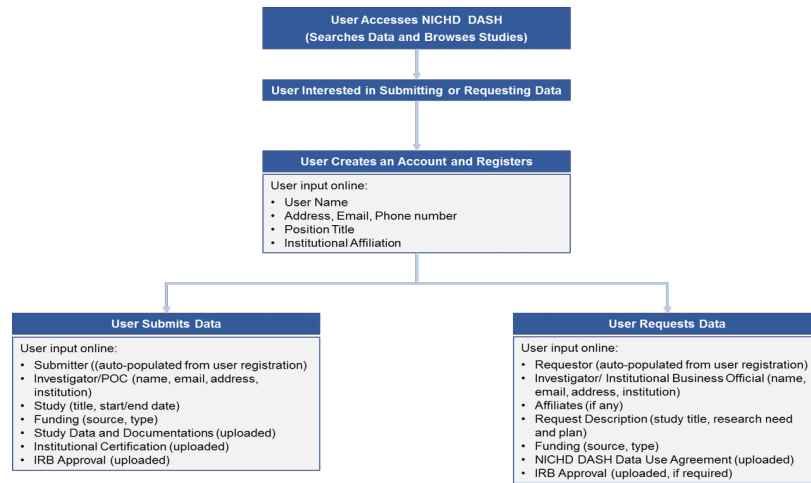
### A.2    Purpose and Use of the Information Collection

As part of the registration, data submission, and data request processes in NICHD DASH, Users are asked to provide specific information online through the NICHD DASH website, as shown in Figure 1.

**Users creating an account** will electronically submit essential information necessary to uniquely identify them in NICHD DASH (*see Attachment A2-1 for screenshots of the Registration webpages*).

**Users submitting research study data** to NICHD DASH will be required to provide information about the study investigator and descriptive information about the study. They will also be required to upload study documentations, data, and an Institutional Certification from the submitting institution stating that the data has been de-identified to the standards set forth in the HHS Regulations for the Protection of Human Subjects and that an Institutional Review Board or Privacy Committee has assessed the proposed data sharing for risks, privacy considerations, and alignment with informed consent (*see Attachment A2-2 for*

**Figure 1: User Information Collection in NICHD DASH**



screenshots of the Data Submission webpages). Users will also be asked to complete a short survey about their experience using DASH after requesting data. The information collected will be reviewed by the NICHD DASH Committee to improve DASH user experience.

**Users creating an account** will electronically submit essential information necessary to uniquely identify them in NICHD DASH (*see Attachment A2-1 for screenshots of the Registration webpages*).

**Users submitting research study data** to NICHD DASH will be required to provide information about the study investigator and descriptive information about the study. They will also be required to upload study documentations, data, and an Institutional Certification from the submitting institution stating that the data has been de-identified to the standards set forth in the HHS Regulations for the Protection of Human Subjects and that an Institutional Review Board or Privacy Committee has assessed the proposed data sharing for risks, privacy considerations, and alignment with informed consent (*see Attachment A2-2 for screenshots of the Data Submission webpages*). Users will also be asked to complete a short survey about their experience using DASH after requesting data. The information collected will be reviewed by the NICHD DASH Committee to improve DASH user experience.

**Users requesting de-identified research study data** from NICHD DASH will be required to provide information about the study investigator and a brief description of the proposed research use of the data requested from NICHD DASH. The investigator and the authorized institutional official will be required to sign a NICHD DASH Data Use Agreement stating that the data recipient will use the data only for the approved research, will not share data with individuals other than those listed in the request, will protect data confidentiality, will not attempt to identify individual participants from whom data were obtained, and will follow appropriate data security protections (*see Attachment A2-3 for screenshots of the Data Request webpages*). The NICHD DASH Data Access Committee will review requests for data to determine whether the proposed research use is scientifically and ethically appropriate and does not conflict with constraints or data use limitations identified by the institutions that submitted the data to NICHD DASH. Users will also be asked to fill out a short survey about their experience using DASH after a submitting a data request. In order to monitor DASH's effectiveness as a platform for data sharing, Data Requesters will be asked to

submit an annual report summarizing research accomplishments, patent applications (or approvals), and any updates to the list of approved Data Users.

The information collected is limited to the essential data required to ensure that the management of Users in NICHD DASH is efficient and the sharing of data among investigators is effective. The primary uses of the information collected from Users by NICHD will be to:

- Communicate with the Users with regards to their data submission or requests
- Monitor data submissions and data requests
- Notify interested recipients of updates to data stored in NICHD DASH
- Help NICHD understand the use of NICHD DASH data by the research community

There are currently no plans to publish information gathered as part of the data access/submission processes because the information will be used only for internal monitoring purposes.

### A.3    Use of Information Technology and Burden Reduction

User information collected in NICHD DASH will be through the web-based portal that enables Users to electronically register for an account, request data access, and submit data. User accounts will be automatically generated. If the User who registers in the system is same as the Submitter or Requestor, the system will auto-populate the User information fields from the registration page minimizing the burden on the User. Similarly, any study information field that recurs in the system will be auto-populated from prior entry. NICHD DASH is designed such that Users will not be asked to enter information more than once in the system.

A  Privacy Impact Assessment (PIA) has been completed for NICHD DASH by the NICHD Privacy Office. NICHD DASH will operate in accordance with existing NIH policies and the Federal Privacy Act to ensure that no sensitive or personally identifiable information, located in federal systems of records is being shared in violation of these policies.

### A.4    Efforts to Identify Duplication and Use of Similar Information

NICHD DASH is primarily a resource for the biomedical research community that includes both NICHD funded and non-funded investigators. Information collected from these Users is not available in any other systems or federal records; hence this data collection is unique.

### A.5    Impact on Small Businesses or Other Small Entities

No small businesses will be involved in this study.

### A.6    Consequences of Collecting the Information Less Frequently

Information will be collected only once from each User for each study submission or request. Once entered into NICHD DASH, the system will store the data and auto-populate whenever the User performs other functions related to the specific study.

The information requested in NICHD DASH electronic forms does not ask Users to generate any new information other than what they already have available and is fundamental to conducting any research study. The information is gathered following a User initiated request and is collected on an as needed basis.

**A.7      Special Circumstances Relating to the Guidelines of 5 CFR 1320.5**

Not Applicable.

**A.8.1    Comments in Response to the Federal Register Notice**

The 60-Day Federal Register notice was published on October 30, 2015 in Vol. 80, pages 66913-4.  No public comments were received.

**A.8.2    Efforts to Consult Outside Agency**

During the planning phase of establishing a data archive, the NICHD DASH Committee conducted an extensive feasibility analysis of over 18 NIH and external data archives to determine if any of the existing archives would be adequate to meet NICHD's data sharing goals. The NICHD DASH Committee consulted experts and viewed demonstrations from many of the archives, including the National Heart Lung and Blood Institute (NHLBI) Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC), the National Children's Study (NCS), NICHD's Biospecimen Repository Access and Data Sharing program (BRADS) and CDC's National Health and Nutrition Examination Survey (NHANES). The feasibility analysis included evaluations of the breadth of research data topics and types, ease of data submission and discovery, policies and governance, system scalability and flexibility, and advanced functionality such as data analytics and linkage to biospecimens.

**A.9      Explanation of Any Payment of Gift to Respondents**

No incentives will be provided to respondents.

**A.10     Assurance of Confidentiality Provided to Respondents**

Data collected in NICHD DASH will be stored and used according to the Federal Privacy Act of 1974. The Federal Privacy Act ensures that no sensitive or personally identifiable information, located in federal systems of records (e.g., Recipient NIH records), is being shared. A system of records is any group of records under the control of a federal agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual.

The information requested from the User seeking to submit or request data in NICHD DASH may be made public in part or in whole for tracking and reporting purposes. A  Privacy Impact Assessment (PIA) has been completed for NICHD DASH by the NICHD Privacy Office. NICHD DASH will operate in accordance with existing NIH policies and the Federal Privacy Act to ensure that no sensitive or personally identifiable information, located in federal systems of records is being shared in violation of these policies.

To ensure that the identities of **research subjects** cannot be readily ascertained from the research study data obtained from NICHD DASH, only research data that are without identifiers and coded will be stored in

NICHD DASH. NICHD will not hold direct identifiers to individuals whose research data are stored within NICHD DASH nor will NICHD have access to the link between the data code and identifiers that may reside with the primary investigators and institutions for particular studies.

**Data submitters** through an Institutional Certification is required to provide assurance that the data has been de-identified to the standards set forth in the HHS Regulations for the Protection of Human Subjects and that an Institutional Review Board or Privacy Committee has assessed the proposed data sharing for risks, privacy considerations, and alignment with informed consent. Research data submitted to NICHD DASH will be stored initially in a staging area until the de-identification status of the research study data is verified by the archive staff and approved by the respective NICHD Division/Branch/Center Chief of the study. Only research study data that have been approved will be shared with the research community.

**Data requestors** through the NICHD DASH Data Use Agreement established between NICHD and the data requestor provides a Privacy Act Notification pursuant to Public Law 93-579, Privacy Act of 1974, 5 U.S.C. Section 552a. Authority for the collection of the information requested from Users comes from the authorities regarding the establishment of the National Institutes of Health, its general authority to conduct and fund research and to provide training assistance, and its general authority to maintain records in connection with these and its other functions (42 U.S.C. 203, 241, 289l-1 and 44 U.S.C. 3101), and Section 301 and 493 of the Public Health Service Act. These records will be maintained in accordance with the Privacy Act System of Record Notice 09-25-0200 (http://oma.od.nih.gov/public/ms/privacy/pafiles/0200.htm) covering "Clinical, Basic and Population-based Research Studies of the National Institutes of Health (NIH), HHS/NIH/OD".

## A.11    Justification for Sensitive Questions

The 'multi-race' category question in the Data Submission Form (Attachment A2-2 Data Submission) is of sensitive nature but it is included because DASH is an archive for storing data from NICHD funded studies that have been completed, including those conducted prior to 1977, when OMB defined new race standards and eliminated the 'multi-race' category. NICHD DASH preserves the originally collected data 'as is' and does not manipulate any of the fields. Therefore, NICHD DASH system will need to reflect race categories that these older studies used.

None of the other information collected during data submission or request in NICHD DASH include questions of a sensitive nature, such as salary, Social Security number, use of alcohol or drugs, sexual behavior and attitudes, religious beliefs, and other matters that are commonly considered private.

### A.12.1  Estimated Annualized Burden Hours

| Form | Number of Respondents | Frequency of Response | Average Time Per Response (in Hours) | Total Annual Burden Hour |
|---|---|---|---|---|
| **Annual Burden Hours Estimate** | | | | |
| User Registration | 120 | 1 | 5/60 | 10 |
| Data Submission | 36 | 1 | 2 | 72 |
| Data Request | 60 | 1 | 1 | 60 |
| **Total** | **120*** | **216*** | | **142** |

*The total number of respondents of 120 include all respondents who are required to complete registration (100) and a subset of those registered respondents who perform data submission (36) or data request (60). The total frequency of response of 216 include the frequency for registration (120*1) as well as for data submissions (36*1) and requests (60*1).

### A.12.2  Annualized Cost to Respondents

| Form | Total Annual Burden Hours | Wage rate | Total Costs |
|---|---|---|---|
| **Total Annual Cost Burden Estimate** | | | |
| User Registration | 10 | $43.35 | $ 433.50 |
| Data Submission | 72 | $43.35 | $ 3121.20 |
| Data Request | 60 | $43.35 | $ 2601.00 |
| **Total Cost** | | | **$ 6155.70** |

Salary/Wage Source: Bureau of Labor Statistics/Occupational Employment and Wages, May 2014: 19-1042 Medical Scientists national estimates: http://www.bls.gov/oes/current/oes191042.htm

### A.13    Estimate of Other Total Annual Cost Burden to Respondents or Record Keepers

There are no additional costs other than the respondents' burden given in A12.

### A.14    Annualized Cost to the Federal Government

The annualized cost to the federal government is $ 14,496.

| Staff | Grade/Step | Salary | % of Effort | Fringe (if applicable) | Total Cost to Gov't |
|---|---|---|---|---|---|
| **Federal Oversight -** Application Lead/ Program Officer | GS-15/ Step 4 | $ 137,494 | 2.0 | | $ 2,846.00 |

| | | $ 116,500 | 10.0 | N/A | $ 11,650.00 |
|---|---|---|---|---|---|
| **Contractor Cost** – Archive Administrator/ Content Analyst | | | | | |
| **Total Cost** | | | | | **$ 14,496.00** |

Salary/Wage Source: Office of Personnel Management 2014 General Schedule Locality Salary Table for various GS-levels; contractor rates based on GSA IT schedules.

### A.15    Explanation for Program Changes or Adjustments

This is a new data collection.

### A.16    Plans for Tabulation and Publication and Project Time Schedule

There is no plan to publish the data collected from Users in NICHD DASH. The data are for the purposes of internal administrative management of NICHD DASH.

### A.17    Reason(s) Display of OMB Expiration Date is Inappropriate

Not applicable.

### A.18    Exceptions to Certification for Paperwork Reduction Act Submissions

The data in NICHD DASH are collected in a manner consistent with the certification statement. No exceptions are requested.

**ATTACHMENTS**

**Attachment A.2-1:  NICHD DASH – User Registration**

**Attachment A2-2:  NICHD DASH – Data Submission**

**Attachment A2-3:  NICHD DASH – Data Request**