Introduction to the NIMH Data Archive (NDA)

Understanding the Basics and Steps Involved

Morgan Justice, Co-I BackInAction

8 August 2024

NIH Collaboratory

Agenda

- NDA overview
- BackInAction context
- Data Sharing Agreements
- Preparing Data for Submission
- Other Important Things

Overview of NIMH Data Archive

- What is the NIMH Data Archive?
  - A repository of research data funded by the National Institute of Mental Health (NIMH).
  - Supports data sharing and reuse to accelerate scientific discovery.
  - Includes data from various research areas like mental health, genetics, and clinical studies.
  - Includes clinical and phenotypic, genomic, neuroimaging, environmental, and biological samples data

BackInAction Context

- NIH data sharing requirements posted after study was funded
- Allowed study sites to vary consenting materials
- Strict privacy requirements due to HIPAA and healthcare systems
- Primary analysis is complete, manuscript under development
  - *Ideally, studies get started with NDA within 6 months of award*
- Sharing analytic datasets and code, but not sharing raw data
- HEAL Initiative study

BackInAction NIMH DA Status

- Preparing data sharing agreement
- Have met with head of NIMH DA, Dan James
- Preparing data for submission
- Have not been assigned a curator

Steps to Submit Data

- Steps for Data Submission
  - Register your study with the NDA.
  - Prepare your data according to NDA guidelines.
  - Submit data using the NDA Data Submission Tool.
  - Ensure compliance with data standards and documentation.

Steps to Submit Data

- Steps for Data Submission
  - **Register your study with the NDA.**
  - Prepare your data according to NDA guidelines.
  - Submit data using the NDA Data Submission Tool.
  - Ensure compliance with data standards and documentation.

Register your study with the NDA: Create an NDA account

- Visit the NDA Website: Go to the official NIMH Data Archive websit https://nda.nih.gov/
- Register for an Account: Click on the registration link and fill out the required information to create your user account. You will need to provide your name, email address, institutional affiliation, and other relevant details.

Register your study with the NDA: Data Submission Agreement (DSA)

- Complete the DSA Form: This form outlines the terms and conditions for using the NDA. It includes details about your research project, the type of data you will be submitting, and how the data will be used.
- Institutional Review Board (IRB) Approval: Obtain IRB approval for your study. This ensures that your research complies with ethical standards and protects the rights and welfare of study participants.
- Submit the DSA: Once completed and signed by your institution, submit the DSA to the NDA through their online portal.

Register your study with the NDA: Data Submission Agreement (DSA)

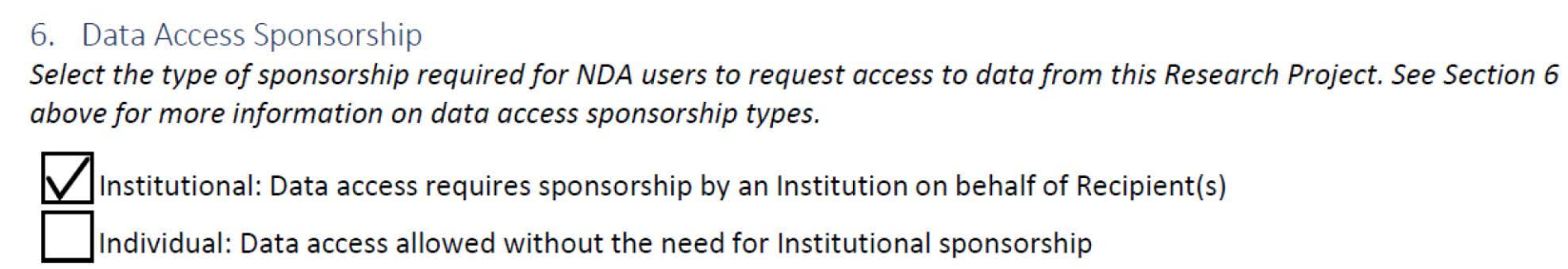
- **Complete the DSA Form: This form outlines the terms and conditions for using the NDA. It includes details about your research project, the type of data you will be submitting, and how the data will be used.**
- Institutional Review Board (IRB) Approval: Obtain IRB approval for your study. This ensures that your research complies with ethical standards and protects the rights and welfare of study participants.
- Submit the DSA: Once completed and signed by your institution, submit the DSA to the NDA through their online portal.

Register your study with the NDA: Data Submission Agreement (DSA)

- Complete the DSA Form: This form outlines the terms and conditions for using the NDA. It includes details about your research project, the type of data you will be submitting, and how the data will be used.
  - Data Access Sponsorship
  - Participant Consent-based Limitations → Extramural Certification Form

Register your study with the NDA: Data Submission Agreement (DSA)

- Complete the DSA Form: This form outlines the terms and conditions for using the NDA. It includes details about your research project, the type of data you will be submitting, and how the data will be used.
  - Data Access Sponsorship
  - Participant Consent-based Limitations → Extramural certification form



6. Data Access Sponsorship

Select the type of sponsorship required for NDA users to request access to data from this Research Project. See Section 6 above for more information on data access sponsorship types.

☑ Institutional: Data access requires sponsorship by an Institution on behalf of Recipient(s)

☐ Individual: Data access allowed without the need for Institutional sponsorship

Register your study with the NDA: Data Submission Agreement (DSA)

- Complete the DSA Form: This form outlines the terms and conditions for using the NDA. It includes details about your research project, the type of data you will be submitting, and how the data will be used.
  - Data Access Sponsorship
  - Participant Consent-based Limitations → Extramural Certification Form

4. Participant Consent

NIMH expects that most data submitted to the NIMH Data Archive will be consented for broad research use. Research Projects submitting data to NDA that have consent-based data use limitations must provide with this DSA an Institutional Certification (https://sharing.nih.gov/genomic-data-sharing-policy/institutional-certifications/completing-an-institutional-certification-form), signed by the Submitter and an Authorized Institutional Business Official, that describes consent-based data use limitations. Submitted data withconsent-based data use limitations will be made available to authorized users in a controlled access Permission Group (https://nda.nih.gov/nda/standard-operating-procedures.html#sop4a).

Check here if data have consent-based limitations on research use ☐

Register your study with the NDA: Data Use Limitations

- Complete the DSA Form: This form outlines the terms and conditions for using the NDA. It includes details about your research project, the type of data you will be submitting, and how the data will be used.
  - Data Access Sponsorship
  - Participant Consent-based Limitations → Extramural Certification Form

Data Use Limitations

General Research Use	GRU	Use of the data is limited only by the terms of the Data Use Certification: these data will be added to the dbGaP Collection.
Health/Medical/Biomedical	HMB	Use of the data is limited to health/medical/biomedical purposes, does not include the study of population origins or ancestry.
Disease-specific [list disease]	DS	Use of the data must be related to the specified disease.
Other		[ENTER CUSTOMIZED TEXT, IF APPLICABLE]

Data Use Limitation Modifiers (Optional)

IRB Approval Required	IRB	Requestor must provide documentation of local IRB approval.
Publication Required	PUB	Requestor agrees to make results of studies using the data available to the larger scientific community.
Collaboration Required	COL	Requestor must provide a letter of collaboration with the primary study investigator(s).
Not-for-profit Use Only	NPU	Use of the data is limited to not-for-profit organizations.
Methods	MDS	Use of the data includes methods development research (e.g., development and testing of software or algorithms).
Genetic Studies Only	GSO	Use of the data is limited to genetic studies only.

Register your study with the NDA: IRB & Consent

- Complete the DSA Form: This form outlines the terms and conditions for using the NDA. It includes details about your research project, the type of data you will be submitting, and how the data will be used.
- **Institutional Review Board (IRB) Approval: Obtain IRB approval for your study. This ensures that your research complies with ethical standards and protects the rights and welfare of study participants.**
- Submit the DSA: Once completed and signed by your institution, submit the DSA to the NDA through their online portal.

Register your study with the NDA: IRB & Consent

- Institutional Review Board (IRB) Approval: Obtain IRB approval for your study. This ensures that your research complies with ethical standards and protects the rights and welfare of study participants.
  - Make sure your consent allows for sharing data in a public data repository

BackInAction Study Example

- Site consent language varied
- 2 sites covered sharing deidentified data in a public repository, 2 sites did not
- Local IRB review
- sIRB review (BIA is in this stage)

Steps to Submit Data

- Steps for Data Submission
  - Register your study with the NDA.
  - **Prepare your data according to NDA guidelines.**
  - Submit data using the NDA Data Submission Tool.
  - Ensure compliance with data standards and documentation.

Data Submission Tools and Resources

- Tools Provided by NDA
  - Data Dictionary (aka Data Structures): Standardizes data elements and definitions. Provides a comprehensive guide to data terms used in NDA.
  - Data Submission Tool: Facilitates secure data upload. Allows for batch uploads and automated data checks.
  - Validation Tools: Ensure data quality and completeness. Includes tools for checking data against predefined standards and identifying errors.

- Support and Training
  - User guides and manuals: Detailed documentation to help users navigate the NDA tools and processes.
  - Training webinars and tutorials: Regular sessions to educate users on data submission and usage.
  - Helpdesk support: Assistance provided by NDA staff for any issues or queries related to data submission and usage.

Data Submission Tools and Resources

- Tools Provided by NDA
  - **Data Dictionary (aka Data Structures): Standardizes data elements and definitions. Provides a comprehensive guide to data terms used in NDA.**
  - Data Submission Tool: Facilitates secure data upload. Allows for batch uploads and automated data checks.
  - Validation Tools: Ensure data quality and completeness. Includes tools for checking data against predefined standards and identifying errors.

- Support and Training
  - User guides and manuals: Detailed documentation to help users navigate the NDA tools and processes.
  - Training webinars and tutorials: Regular sessions to educate users on data submission and usage.
  - Helpdesk support: Assistance provided by NDA staff for any issues or queries related to data submission and usage.

Data Submission Key Concepts

- Data structures
  - Required fields
  - Aligning with existing data structures v. creating a new one
  - HEAL Common Data Elements
- Data file formats
- GUID and pseudoGUID
- Data curators
- Supporting information
- Are your measures already programmed or collected?

Data Submission Key Concepts: Data Structures

- **Data structures**
  - **Required fields**
  - Aligning with existing data structures v. creating a new one
  - HEAL Common Data Elements
- Data file formats
- GUID and pseudoGUID
- Data curators
- Supporting information
- Are your measures already programmed or collected?

Data Submission Key Concepts: Data Structures

Data Structure

Patient Reported Outcome Measurement Information System (PROMIS) -57 Version 01

Short Name: promis5701

Data Type: Clinical Assessments **Categories:** Behavior

PROMIS-57 is a questionnaire for self-reporting different aspects of physical, mental, and social health in adults. There are sections for physical function, anxiety, depression, fatigue, sleep problems, social participation and pain measurement.

QUERY THIS STRUCTURE

DOWNLOADS

Definition

Submission Template

A Collection's submission must be cumulative.

ELEMENTS | VIEW CHANGE HISTORY

Filter Search

+ Add Column

Name, Type, Required, Description, Value...

62 **Elements**

NAME	TYPE	REQUIRED	DESCRIPTION	VALUE RANGE	NOTES
subjectkey	GUID	Required	The NDAR Global Unique Identifier (GUID) for research subject	NDAR*	
src_subject_id	String	Required	Subject ID how it's defined in lab/project		
interview_date	Date	Required	Date on which the interview/genetic test/sampling/imaging/biospecimen was completed. MM/DD/YYYY		
interview_age	Integer	Required	Age in months at the time of the interview/test/sampling/imaging.	0::1440	Age is rounded to chronological month. If the research participant is 15-days-old at time of interview, the appropriate value would be 0 months. If the participant is 16-days-old, the value would be 1 month.
gender	String	Required	Sex of subject at birth	M;F; O; NR	M = Male; F = Female; O=Other; NR = Not reported

Data Submission Key Concepts: Data Structures: Required Fields

NAME	TYPE	REQUIRED	DESCRIPTION
subjectkey	GUID	Required	The NDAR Global Unique Identifier (GUID) for research subject
src_subject_id	String	Required	Subject ID how it's defined in lab/project
interview_date	Date	Required	Date on which the interview/genetic test/sampling/imaging/biospecimen was completed. MM/DD/YYYY
interview_age	Integer	Required	Age in months at the time of the interview/test/sampling/imaging.
gender	String	Required	Sex of subject at birth

Data Submission Key Concepts: Data Structures: Required Fields

NAME	TYPE	REQUIRED	DESCRIPTION
subjectkey	GUID	Required	The NDAR Global Unique Identifier (GUID) for research subject
src_subject_id	String	Required	Subject ID how it's defined in lab/project
interview_date	Date	Required	Date on which the interview/genetic test/sampling/imaging/biospecimen was completed. MM/DD/YYYY
interview_age	Integer	Required	Age in months at the time of the interview/test/sampling/imaging.
gender	String	Required	Sex of subject at birth

Exceptions to Data Submission

Data collection date (interview_date) and age in months (interview_age) are required data fields in all NDA data structures. However, institutional IRBs may restrict submission of date and age variables into NDA collections. In this case, collections should follow the NDA data masking methodology, where date or age are modified by a persistent "offset" number that is generated for each subject, securely stored by the collection administrators, and not shared outside of the study. Dates may also be masked by using a common month or calendar day for all entries in a dataset. This common methodology will support appropriate secondary use of the data.

Data Submission Key Concepts: Data Structures: Required Fields

NAME	TYPE	REQUIRED	DESCRIPTION
subjectkey	GUID	Required	The NDAR Global Unique Identifier (GUID) for research subject
src_subject_id	String	Required	Subject ID how it's defined in lab/project
interview_date	Date	Required	Date on which the interview/genetic test/sampling/imaging/biospecimen was completed. MM/DD/YYYY
interview_age	Integer	Required	Age in months at the time of the interview/test/sampling/imaging.
gender	String	Required	Sex of subject at birth

Dan James from NIMH DA also told us that we could set these two variables to missing if our IRB would not allow any masked dates to be shared

Exceptions to Data Submission

Data collection date (interview_date) and age in months (interview_age) are required data fields in all NDA data structures. However, institutional IRBs may restrict submission of date and age variables into NDA collections. In this case, collections should follow the NDA data masking methodology, where date or age are modified by a persistent "offset" number that is generated for each subject, securely stored by the collection administrators, and not shared outside of the study. Dates may also be masked by using a common month or calendar day for all entries in a dataset. This common methodology will support appropriate secondary use of the data.

Data Submission Key Concepts: Data Structures

- **Data structures**
  - Required fields
  - **Aligning with existing data structures v. creating a new one**
  - HEAL Common Data Elements
- Data file formats
- GUID and pseudoGUID
- Data curators
- Supporting information
- Are your measures already programmed or collected?

Data Submission Key Concepts:
Data Structures: Existing Structures

NAME	TYPE	REQUIRED	DESCRIPTION	VALUE RANGE	NOTES
rpd_01	Integer	Recommended	I stay at home most of the time because of my pain.	0;1;99	0= No; 1= Yes; 99= NA
rpd_02	Integer	Recommended	I change position frequently to try to get my pain comfortable.	0;1;99	0= No; 1= Yes; 99= NA
rpd_03	Integer	Recommended	I walk more slowly than usual because of my pain.	0;1;99	0= No; 1= Yes; 99= NA
rpd_04	Integer	Recommended	Because of my pain, I am not doing any of the jobs that I usually do around the house.	0;1;99	0= No; 1= Yes; 99= NA
rpd_05	Integer	Recommended	Because of my pain, I use a handrail to get upstairs.	0;1;99	0= No; 1= Yes; 99= NA
rpd_06	Integer	Recommended	Because of my pain, I lie down to rest more often.	0;1;99	0= No; 1= Yes; 99= NA

ROLAND (all Roland items will be preceded by ROLAND_)

When your back hurts, you may or may not find it difficult to do some of the things you norma do. The first questions include statements that people sometimes use to describe themselv when they have back pain. Please think of the past week. Answer "yes" if the statement describ you and "no" if it does not. In the past week...

1. Roland Morris Disability Questionnaire	NO	YES	DK	REF
A. I stay at home most of the time because of the pain in my back.	0	1	8	9
B. I change position frequently to try and make my back comfortable.	0	1	8	9
C. I walk more slowly than usual because of the pain in my back.	0	1	8	9
D. Because of the pain in my back, I am not doing any of the jobs that I usually do around the house.	0	1	8	9
E. Because of the pain in my back, I use a handrail to get upstairs.	0	1	8	9
F. Because of the pain in my back, I lie down to rest more often.	0	1	8	9

Steps:

1. Search for your instrument/data element
2. Compare description with question wording
3. Compare values and labels/scales

Data Submission Key Concepts: Data Structures: Existing Structures

NAME	TYPE	REQUIRED	DESCRIPTION	VALUE RANGE	NOTES
rpd_01	Integer	Recommended	I stay at home most of the time because of my pain.	0;1;99	0= No; 1= Yes; 99= NA
rpd_02	Integer	Recommended	I change position frequently to try to get my pain comfortable.	0;1;99	0= No; 1= Yes; 99= NA
rpd_03	Integer	Recommended	I walk more slowly than usual because of my pain.	0;1;99	0= No; 1= Yes; 99= NA
rpd_04	Integer	Recommended	Because of my pain, I am not doing any of the jobs that I usually do around the house.	0;1;99	0= No; 1= Yes; 99= NA
rpd_05	Integer	Recommended	Because of my pain, I use a handrail to get upstairs.	0;1;99	0= No; 1= Yes; 99= NA
rpd_06	Integer	Recommended	Because of my pain, I lie down to rest more often.	0;1;99	0= No; 1= Yes; 99= NA

ROLAND (all Roland items will be preceded by ROLAND_)

When your back hurts, you may or may not find it difficult to do some of the things you norma do. The first questions include statements that people sometimes use to describe themselv when they have back pain. Please think of the past week. Answer "yes" if the statement describ you and "no" if it does not. In the past week...

1. Roland Morris Disability Questionnaire	NO	YES	DK	REF
A. I stay at home most of the time because of the pain in my back.	0	1	8	9
B. I change position frequently to try and make my back comfortable.	0	1	8	9
C. I walk more slowly than usual because of the pain in my back.	0	1	8	9
D. Because of the pain in my back, I am not doing any of the jobs that I usually do around the house.	0	1	8	9
E. Because of the pain in my back, I use a handrail to get upstairs.	0	1	8	9
F. Because of the pain in my back, I lie down to rest more often.	0	1	8	9

If using an existing data structure, variable names & missing data values will need to be aligned if different from each other

Data Submission Key Concepts: Data Structures: New Structures

Two scenarios that require new structures:

1. Your instrument or data element doesn't have a corresponding data structure
2. The existing data structure scale or categories are either incorrect or do not match your data element

Data Submission Key Concepts: Data Structures: New Structures

Two scenarios that require new structures:

1. **Your instrument or data element doesn't have a corresponding data structure**
2. The existing data structure scale or categories are either incorrect or do not match your data element

Many EHR data elements will require a new data structure to be defined unless they are a common PROM, e.g, PHQ or GAD

Data Submission Key Concepts: Data Structures: New Structures

Two scenarios that require new structures:

1. The instrument or data element doesn't have a corresponding data structure
2. **The existing data structure scale or categories are either incorrect or do not match your data element**

promis57_pfa11	Integer	Recommended	Are you able to do chores such as vacuuming or yard work?	1::6; -999	1=Without difficulty; 2=With a little difficulty; 3=With some difficulty; 4=With much difficulty; 5=Unable to do; 6=Rather not answer; -999 = Missing
promis57_pfa21	Integer	Recommended	Are you able to go up and down stairs at a normal pace?	1::6; -999	1=Without difficulty; 2=With a little difficulty; 3=With some difficulty; 4=With much difficulty; 5=Unable to do; 6=Rather not answer; -999 = Missing

Data Submission Key Concepts: Data Structures: New Structures

Two scenarios that require new structures:

1. Your instrument or data element doesn't have a corresponding data structure
2. **The existing data structure scale or categories are either incorrect or do not match your data element**

promis57_pfa11	Integer	Recommended	Are you able to do chores such as vacuuming or yard work?	1::6; -999	1=Without difficulty; 2=With a little difficulty; 3=With some difficulty; 4=With much difficulty; 5=Unable to do; 6=Rather not answer; -999 = Missing
promis57_pfa21	Integer	Recommended	Are you able to go up and down stairs at a normal pace?	1::6; -999	1=Without difficulty; 2=With a little difficulty; 3=With some difficulty; 4=With much difficulty; 5=Unable to do; 6=Rather not answer; -999 = Missing

	Without any difficulty	With a little difficulty	With some difficulty	With much difficulty	Unable to do	DK	REF
PROMIS_SF6B_1 Are you able to do chores such as vacuuming or yard work?	5	4	3	2	1	8	9
PROMIS_SF6B_2 Are you able to go up and down stairs at a normal pace?	5	4	3	2	1	8	9

Data Submission Key Concepts: Data Structures: New Structures

Two scenarios that require new structures:

1. Your instrument or data element doesn't have a corresponding data structure
2. **The existing data structure scale or categories are either incorrect or do not match your data element**

promis57_pfa11	Integer	Recommended	Are you able to do chores such as vacuuming or yard work?	1::6; -999	1=Without difficulty; 2=With a little difficulty; 3=With some difficulty; 4=With much difficulty; 5=Unable to do; 6=Rather not answer; -999 = Missing
promis57_pfa21	Integer	Recommended	Are you able to go up and down stairs at a normal pace?	1::6; -999	1=Without difficulty; 2=With a little difficulty; 3=With some difficulty; 4=With much difficulty; 5=Unable to do; 6=Rather not answer; -999 = Missing

	Without any difficulty	With a little difficulty	With some difficulty	With much difficulty	Unable to do	DK	REF
PROMIS_SF6B_1 Are you able to do chores such as vacuuming or yard work?	5	4	3	2	1	8	9
PROMIS_SF6B_2 Are you able to go up and down stairs at a normal pace?	5	4	3	2	1	8	9

Data Submission Key Concepts: Data Structures

- **Data structures**
  - Required fields
  - Aligning with existing data structures v. creating a new one
  - **HEAL Common Data Elements**
- Data file formats
- GUID and pseudoGUID
- Data curators
- Supporting information
- Are your measures already programmed or collected?

Data Submission Key Concepts: Data Structures: HEAL CDEs

- HEAL Data Stewards working with NIMH DA to create data structures that reflect the HEAL CDEs
- However, timeline for having these data structures ready is not clear

BackInAction (and RESOLVE) Study Example

- We are aligning our variable names with existing NIMH DA data structures despite having used HEAL CDEs

Data Submission Key Concepts: Data Structures

Important Reminders:

- Compare your instruments against the NIMH DA data structures very closely
- Misaligned variable names or missing data values do not require new data structures; only if the scale or categories are different or your data element isn't available on NDA

Data Submission Key Concepts: Data File Formats

- Data structures
  - Required fields
  - Aligning with existing data structures v. creating a new one
  - HEAL Common Data Elements
- **Data file formats**
- GUID and pseudoGUID
- Data curators
- Supporting information
- Are your measures already programmed or collected?

Data Submission Key Concepts: Data File Formats

- NIMH DA requests each data structure, e.g., instrument, be uploaded as separate CSV, TSV, JSON, XML, imaging data formatted, or genomic data formatted files
- However, they also allow flexibility in this and can accept a single dataset

Data Submission Key Concepts: GUIDs and pseudoGUIDs

- Data structures
  - Required fields
  - Aligning with existing data structures v. creating a new one
  - HEAL Common Data Elements
- Data file formats
- **GUID and pseudoGUID**
- Data curators
- Supporting information
- Are your measures already programmed or collected?

Data Submission Key Concepts: GUIDs and pseudoGUIDs

NAME	TYPE	REQUIRED	DESCRIPTION
subjectkey	GUID	Required	The NDAR Global Unique Identifier (GUID) for research subject
src_subject_id	String	Required	Subject ID how it's defined in lab/project
interview_date	Date	Required	Date on which the interview/genetic test/sampling/imaging/biospecimen was completed. MM/DD/YYYY
interview_age	Integer	Required	Age in months at the time of the interview/test/sampling/imaging.
gender	String	Required	Sex of subject at birth

Data Submission Key Concepts: GUIDs and pseudoGUIDs

Global Unique Identifier (GUIDs)

- https://nda.nih.gov/nda/using-the-nda-guid
- Alphanumeric code that is created by the NIMH Data Archive (NDA) GUID Tool and used as an identifier for a research participant
- The GUID itself is not personally identifiable information or protected health information.
- However, there are some very important elements to understand about the GUID
  - How GUIDs are created
  - What GUIDs allow
  - Alternatives to the GUID, i.e., pseudoGUIDs

Data Submission Key Concepts: How GUIDs are Created

How does the NDA GUID Tool work?

1. An authorized member of the research project team (user) downloads the NDA GUID Tool.
2. The user enters participant PII into the tool.
3. In a local computation, the tool generates a series of one-way hash codes based on the PII entered, without the PII ever leaving your computer.
4. The hash codes are encrypted and securely sent to the GUID system at NDA.
5. If the hash codes match an existing hash code, the GUID associated with that hash code is sent back to the researcher.
6. If the hash codes do not match an existing hash code in the NDA GUID system, a new GUID is created and sent back.

Data Submission Key Concepts: How GUIDs are Created

How does the NDA GUID Tool work?

1. An authorized member of the research project team (user) downloads the NDA GUID Tool.
2. The user enters participant PII into the tool.
3. In a local computation, the tool generates a series of one-way hash codes based on the PII entered, without the PII ever leaving your computer.
4. The hash codes are encrypted and securely sent to the GUID system at NDA.
5. If the hash codes match an existing hash code, the GUID associated with that hash code is sent back to the researcher.
6. If the hash codes do not match an existing hash code in the NDA GUID system, a new GUID is created and sent back.

Data Submission Key Concepts: How GUIDs are Created

PII Required to Create the GUID (created locally)

- First Name
- Middle Name
- Last Name
- Sex
- Date of Birth
- City/Municipality of Birth

All information is required to be recorded exactly as it appears on the *birth certificate*

Data Submission Key Concepts: What GUIDs Allow

With the NDA GUID Tool, the same participant information will return the same GUID whenever or wherever it is entered. This allows NDA to link participant data records across time and locations, without ever receiving identifying information. This system has the following “advantages” (quotations added):

- No PII ever leaves your computer.
- There is nothing about a GUID that would allow someone to infer the identity of the individual to whom it belongs.
- The same individual's information will result in the same GUID across time, location, and research study. This allows authorized researchers to match shared data from that participant regardless of source, without ever sharing or viewing PII.

Data Submission Key Concepts: What GUIDs Allow

With the NDA GUID Tool, the same participant information will return the same GUID whenever or wherever it is entered. This allows NDA to link participant data records across time and locations, without ever receiving identifying information. This system has the following “advantages” (quotations added):

- No PII ever leaves your computer.
- There is nothing about a GUID that would allow someone to infer the identity of the individual to whom it belongs.
- The same individual's information will result in the same GUID across time, location, and research study. This allows authorized researchers to match shared data from that participant regardless of source, without ever sharing or viewing PII.

Data Submission Key Concepts: What GUIDs Allow

With the NDA GUID Tool, the same participant information will return the same GUID whenever or wherever it is entered. This allows NDA to link participant data records across time and locations, without ever receiving identifying information. This system has the following “advantages” (quotations added):

- No PII ever leaves your computer.
- There is nothing about a GUID that would allow someone to infer the identity of the individual to whom it belongs.
- The same individual's information will result in the same GUID across time, location, and research study. This allows authorized researchers to match shared data from that participant regardless of source, without ever sharing or viewing PII.

Data Submission Key Concepts: Alternative to the GUID - pseudoGUIDs

- If any of the information necessary to create a GUID is missing, a pseudoGUID can be created. This is a random ID that can be used as a placeholder where this information is not available, and "promoted" to a real GUID when the information is obtained at a future date.

- To request pseudoGUIDs, users must have both NDA GUID Tool access and must submit a justification request to the NDA Help Desk explaining why GUIDs are not able to be generated.

Data Submission Key Concepts: Alternative to the GUID - pseudoGUIDs

- If any of the information necessary to create a GUID is missing, a pseudoGUID can be created. This is a random ID that can be used as a placeholder where this information is not available, and "promoted" to a real GUID when the information is obtained at a future date.
  - pseudoGUIDs do not have to be “promoted” and can remain the primary participant ID forever

- To request pseudoGUIDs, users must have both NDA GUID Tool access and must submit a justification request to the NDA Help Desk explaining why GUIDs are not able to be generated.

Data Submission Key Concepts: Alternative to the GUID - pseudoGUIDs

- If any of the information necessary to create a GUID is missing, a pseudoGUID can be created. This is a random ID that can be used as a placeholder where this information is not available, and "promoted" to a real GUID when the information is obtained at a future date.

- To request pseudoGUIDs, users must have both NDA GUID Tool access and must submit a justification request to the NDA Help Desk explaining why GUIDs are not able to be generated.
  - Per Dan James, the NDA GUID Tool is not required for pseudoGUIDs. Users simply email the NDA Help Desk with a justification request and the number identifiers needed and the Help Desk sends back a list of pseudoGUIDs.

Data Submission Key Concepts: Data Curators

- Data structures
  - Required fields
  - Aligning with existing data structures v. creating a new one
  - HEAL Common Data Elements
- Data file formats
- GUID and pseudoGUID
- **Data curators**
- Supporting information
- Are your measures already programmed or collected?

Data Submission Key Concepts: Data Curators

- Partner with you to get your data ready to submit
- Work in 2-week sprints
- Primarily communicate via email
- Pace of work depends on study team responsiveness
- 5 or 6 curators supporting >3,000 studies

Data Submission Key Concepts: Supporting Information

- Data structures
  - Required fields
  - Aligning with existing data structures v. creating a new one
  - HEAL Common Data Elements
- Data file formats
- GUID and pseudoGUID
- Data curators
- **Supporting information**
- Are your measures already programmed or collected?

Data Submission Key Concepts: Supporting Information

Supporting Documentation is expected to be submitted to the Research Project's NDA Collection and shared prior to the end of the Research Project.

- Research protocol(s)
- Questionnaire(s)
- Study manuals
- Clinical Trial protocol(s)
- Manual of Procedures and Case Report Forms
- Statistical analysis plans (SAPs)
- Code for analyses

Data Submission Key Concepts: Timing

- Data structures
  - Required fields
  - Aligning with existing data structures v. creating a new one
  - HEAL Common Data Elements
- Data file formats
- GUID and pseudoGUID
- Data curators
- Supporting information
- **Are your measures already programmed or collected?**

Data Submission Key Concepts: Timing

- Study was just funded
  - Work with NIMH DA to align variable names, formats/values before you define variables
- Data elements programmed or collected
  - Incorporate NIMH DA structures into your analytic code and data cleaning to align variable names
- Analyses complete
  - Update variable names, missing values and document in the code where this occurs

Other Important Things: Data Sharing Schedule

- Unless a specific schedule for data sharing has been defined in advance, data will be shared with authorized users upon publication (via an NDA Study) or 1-2 years after the grant end date specified on the first Notice of Award

Other Important Things: Data Embargo

- Unless a specific schedule for data sharing has been defined in advance, data will be shared with authorized users upon publication (via an NDA Study) or 1-2 years after the grant end date specified on the first Notice of Award
- Data embargoes are allowed, typically for 1-year, but extensions are common

Other Important Things: Acknowledging NIMH DA

- Unless a specific schedule for data sharing has been defined in advance, data will be shared with authorized users upon publication (via an NDA Study) or 1-2 years after the grant end date specified on the first Notice of Award
- Data embargoes are allowed, typically for 1-year, but extensions are common
- Submitters agree to acknowledge the appropriate NIMH Data Archive data repository and the relevant Digital Object Identifier(s) (DOI), which will be minted upon NIMH Data Archive Study creation, in any and all oral and written presentations, disclosures, and publications (including abstracts, as space allows) resulting from any and all analyses of data

Getting Started with NIMH DA

- First Steps
  - Visit the NDA website
  - Register and review the Data Sharing Agreement
  - Explore available data structures

- Contact Information
  - NDA Help Desk: NDAHelp@mail.nih.gov.
  - Further webinars and tutorials

Q&A

Morgan.J.Fuoco@kp.org

Data Submission Key Concepts: How GUIDs are Created

How does the NDA GUID Tool work?

1. An authorized member of the research project team (user) downloads the NDA GUID Tool.
2. The user enters participant PII into the tool.
3. In a local computation, the tool generates a series of one-way hash codes based on the PII entered, without the PII ever leaving your computer.
4. The hash codes are encrypted and securely sent to the GUID system at NDA.
5. If the hash codes match an existing hash code, the GUID associated with that hash code is sent back to the researcher. The GUID is an alphanumeric code that is randomly and persistently linked to the hash codes within the secure NDA GUID system and cannot be traced back to the PII entered by the research project team member.
6. If the hash codes do not match an existing hash code in the NDA GUID system, a new GUID is created and sent back.