Data Submission Office Hours

Webinar

April 2023
Purpose

• **Purpose:** Learn the general NDA Data Submission process during a Submission Cycle

• **Agenda:**
  - Who, what, when and how to submit data
  - Prerequisites to data submission
  - How to check if your Data Structures are ready for submission
  - Submission Templates
  - Mandatory Data Structures
  - Submission Exemptions
Who should be submitting data & when?

NDA expects you to submit data biannually by deadlines:

January 15 & July 15

1. Your NDA Collection is in the NIMH Data Archive data repository
2. Your Collection phase is in Pre-Enrollment, Enrolling or Funding Completed – Data Expected
3. The Data Expected list has an initial submission date set to the deadline date or earlier
Who should be submitting data & when?

NDA expects you to submit data biannually by deadlines:

**April 1 & October 1**

1. Your NDA Collection is in the NIAAA Data Archive data repository

2. Your Collection phase is in Pre-Enrollment, Enrolling or Funding Completed – Data Expected

3. The Data Expected list has an initial submission date set to the deadline date or earlier

---

**Data Expected List: Mandatory Data Structures**

These data structures are mandatory for your NDA Collection. Please update for the entire study.

<table>
<thead>
<tr>
<th>Data Expected</th>
<th>Targeted Enrollment</th>
<th>Initial Submission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Subject and Pedigree</td>
<td>46</td>
<td>10/01/2021</td>
</tr>
</tbody>
</table>

**Data Expected List: Data Structures per Research Aims**

---

**Collection State: Shared**

**Collection Type: None**

**Collection Title:** Influence of synthetic sex hormones

**Collection Investigators:** Emma Childs

**Collection Description:** There is ample preclinical evidence to contribute to the enhanced vulnerability

**Data Repository:** NIAAA Data Archive

**Permission Group:** NIAAA Data Archive

**Collection Creation Date:** 08/25/2020

**Collection Phase:** Enrolling

**Collection State:** Shared

**Collection Type:** Default [ ] Adult [ ] Child [ ]
What do I need to start submitting data?

- Data Submission Agreement (DSA)
- Data Expected List
- Permissions to Submit
- Global Unique Identifiers (GUIDs)
Data Submission Agreement (DSA)

- Submit through the eDSA Tool
- Follow the instructions in Welcome to Data Sharing
- Access to your NDA Collection & Data Expected list
Data Expected List

- List of all Data Structures you’ll submit data for
- Admin permissions required to edit
Permissions to Submit

- Submission permissions required to submit data
- Your PI manages permissions via the Permissions tab
Global Unique Identifiers (GUIDs)

- Required for all data submissions
- Access to NDA GUID Tool requested via the NDA Help Desk
- Submission permissions required
Global Unique Identifier (GUID)

Alphanumeric code used as an identifier for every research participant (ex. NDARRF67XH1)

What information do I need to create a GUID?

• First, middle and last name
• Sex at birth
• Date of birth
• City/Municipality of Birth
• *Recorded as it appears on the birth certificate

The NDA Data Sharing Terms and Conditions include the expectation that you will collect the data necessary to create real GUIDs.
pseudoGUIDs can be requested in place of GUIDs by emailing the NDA Help Desk:

- Detailed justification for the use of pseudoGUIDs
- # of pseudoGUIDs needed
- Whether pseudoGUIDs will be needed for more than one submission cycle

What if I obtain the information needed to create a GUID, but I’ve already used a pseudoGUID?

Promote your pseudoGUID using the NDA GUID Tool linking the pseudoGUID to a real GUID
What do I need to start submitting data?

**Data Submission Agreement (DSA)**
- Submit through the eDSA Tool
- Follow the instructions in Welcome to Data Sharing
- Access to your NDA Collection & Data Expected list

**Data Expected List**
- List of all Data Structures you’ll submit data for
- [NDA Data Dictionary](#)
- Admin permissions required to edit

**Permissions to Submit**
- Submission permissions required to submit data
- Your PI manages permissions via the Permissions tab

**Global Unique Identifiers (GUIDs)**
- Required for all data submissions
- Access to NDA GUID Tool requested via the NDA Help Desk
- Submission permissions required
Data Submission

What data should be submitted?

• Please submit all data you have collected.
• You **DO NOT** need to submit data for all the Data Structures.

Should data be submitted cumulatively?

• Clinical and phenotypic data → **cumulative** submissions
  • Your first submission includes 5 subjects at Timepoint A
  • You collect 10 new subjects at Timepoint A & original 5 subjects at Timepoint B
  • Include the 5 subjects Timepoint A & B + 10 new subjects for Timepoint A

• Imaging, EEG, Omics data → **non-cumulative submissions**
Should I submit any data if NDA is still working on your Data Structures?

• Yes!

• You should submit data for Data Structures that are approved and are ready for submission!
How do I know if my Data Structure is ready?

Check the information ("i") icon next to the data structure names via your Data Expected tab!

Structure not yet defined: Data curation is still preparing your Data Structure

If the pop-up provides the title and short name of your data structure: Your data structure is ready for submission!
Download Your Submission Template
Submission Template

• Do not change the first two rows.
• You can hide columns. Do not delete any columns.
• Each row = 1 subject at 1 timepoint
  • Example: 1 subject at 3 timepoints = 3 rows
# Submission Template

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndar_subject</td>
<td>subjectkey</td>
<td>src_subject_id</td>
<td>interview_date</td>
<td>interview_age</td>
<td>sex</td>
<td>race</td>
<td>ethnic_group</td>
<td>phenotype</td>
</tr>
</tbody>
</table>

## Definition

**subjectkey**
- **DataType**: GUID
- **Size**: 20
- **Condition**: Required
- **ElementDescription**: The NDAR Global Unique Identifier (GUID) for research subjects

**src_subject_id**
- **DataType**: String
- **Size**: 20
- **Condition**: Required
- **ElementDescription**: Subject ID how it's defined in lab/project

**interview_date**
- **DataType**: Date
- **Size**: 20
- **Condition**: Required
- **ElementDescription**: Date on which the interview/genetic test/sampling

**interview_age**
- **DataType**: Integer
- **Size**: 20
- **Condition**: Required
- **ElementDescription**: Age in months at the time of the interview/test/sampling

**sex**
- **DataType**: String
- **Size**: 20
- **Condition**: Required
- **ElementDescription**: Sex of subject at birth

**race**
- **DataType**: String
- **Size**: 20
- **Condition**: Required
- **ElementDescription**: Race of study subject

**ethnic_group**
- **DataType**: String
- **Size**: 20
- **Condition**: Recommended
- **ElementDescription**: Ethnic group

**phenotype**
- **DataType**: String
- **Size**: 20
- **Condition**: Required
- **ElementDescription**: Phenotype/diagnosis for the subject

**phenotype_description**
- **DataType**: String
- **Size**: 4000
- **Condition**: Required
- **ElementDescription**: Description of the phenotype for the subject
Required Data Elements

- Required data elements must have a value.
- Conditional data elements should have a value if applicable.
- Recommended and optional data elements can be left blank if not applicable.
### Required Data Elements

- If you did not collect a required data element, review `ValueRange & Notes` for N/A or null values.

<table>
<thead>
<tr>
<th>ElementName</th>
<th>Size</th>
<th>Required</th>
<th>ValueRange</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex</td>
<td>20</td>
<td>Required</td>
<td>M; F; O; NR</td>
<td>M = Male; F = Female; O = Other; NR = Not reported</td>
</tr>
<tr>
<td>zygosity</td>
<td>255</td>
<td>Conditional</td>
<td>monozygous; dizygous; trizygotic; 999</td>
<td>999 = unknown</td>
</tr>
<tr>
<td>reg_edu</td>
<td></td>
<td>Recommended</td>
<td>1:34; 999</td>
<td>999 = Unknown</td>
</tr>
<tr>
<td>brain_injury</td>
<td></td>
<td>Recommended</td>
<td>1:3</td>
<td>3 = NA</td>
</tr>
</tbody>
</table>
### 5 Required Data Elements in All Structures

<table>
<thead>
<tr>
<th>ELEMENT NAME</th>
<th>DATA TYPE</th>
<th>SIZE</th>
<th>REQUIRED</th>
<th>DESCRIPTION</th>
</tr>
</thead>
</table>


How do I submit data?

Via the **Validation and Upload Tool**!

- Submission permissions required
- HTML Validation and Upload Tool
  - Cumulative data (e.g. clinical and phenotypic data)
  - [Tutorial for the HTML Validation and Upload Tool](#)
- Python Client Validation and Upload Tool
  - Non-cumulative imaging files
  - Files over 2.5 GB

*You **must** use the Submission Templates from the NDA Data Dictionary and/or be provided by a Data Curator to submit your data.*
Validation and Upload Tool

- Use it as your personal NDA data manager
- Warnings vs Errors
  - Errors **must** be resolved
  - Users can submit with warnings present
- Need help? Click “Request Help”

<table>
<thead>
<tr>
<th>File Name</th>
<th>Short Name</th>
<th>Status</th>
<th>Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>test_data.csv</td>
<td></td>
<td>Complete with Errors</td>
<td>1</td>
</tr>
</tbody>
</table>

1 file(s) must be valid before submission. Choose a file above to view errors and warnings.
What if none of your Structures are ready?

• If you are waiting on NDA to complete your data structures where you have ready-to-submit data, don’t worry!

• There will always be at least one ready-to-submit data structure.
Mandatory Data Structures

Research Subject & Pedigree (RS&P)
- ndar_subject01
- genomics_subject02 (omics only)

NIH Common Data Elements (CDE)
(NOT-MH-20-067)
Mandatory Data Structures

- **Update** the targeted enrollment number.
- **Submit** data cumulatively.

![Data Expected vs Targeted Enrollment Table]

<table>
<thead>
<tr>
<th>Data Expected</th>
<th>Targeted Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Subject and Pedigree</td>
<td>1,000</td>
</tr>
<tr>
<td>NIMH CDE - GAD-7</td>
<td>1,000</td>
</tr>
<tr>
<td>NIMH CDE - PHQ-9</td>
<td>1,000</td>
</tr>
<tr>
<td>NIMH CDE - DSM-5 Crosscutting Adults</td>
<td>1,000</td>
</tr>
<tr>
<td>NIMH CDE - WHODAS 2.0</td>
<td>1,000</td>
</tr>
</tbody>
</table>
Don’t have any data to submit / Structures not ready?

Submit a submission exemption via the Data Expected tab of your NDA Collection!

Types of Exemptions

A. **No New Data This Period**: No new data to submit

B. **Deadline Extension**: Data submission expected after the submission deadline

C. **No New Data - Change Phase to Data Analysis**: Data submission is completed for the entire grant
If QA errors were identified from a previous submission cycle, do I still submit data?

- Please **correct the QA errors** and resubmit your data before submitting any new data.
- Follow the instructions in your QA results notification.
More Questions?

If you have further questions or would like a copy of the slide deck, please email the NDA Help Desk at NDAHelp@mail.nih.gov.

All information from the presentation and more can be found on our Submission Cycle FAQ: https://nda.nih.gov/nda/submission-cycle-faq.html.

Thank you for joining!