

Guidance for Preparing REDCap Data for NIMH Data Archive Submission

Purpose

Formatting differences between REDCap and the NIMH Data Archive require adaptation of data stored in REDCap databases. This document provides guidance on the types of data transformations that are often necessary to prepare data collected in REDCap for submission to the NIMH Data Archive (NDA).

Definitions

Data Structure: A Data Structure is an organizing construct in the NDA data dictionary and is a table in the NDA database. Typically, it is a published measure (e.g. Mullen, Parental Stress Index), but may also be a definition for submission of a type of data (imaging or genomics) or a resultset from a computational pipeline (see [Evaluated Data](#) category). A data structure provides a tabular resultset of a subject's item level data. Each data structure in the NDA will contain the GUID, age in months, sex, date of assessment taken (e.g. scan, measure given or blood draw date), and the project's source subject identifier.

Data Element: A Data Element is an individual item (variable/field) within a Data Structure that represents a question/field on an assessment or a discrete data point. The NDA data curation team maps identical Data Elements across Data Structures as part of the NDA data harmonization approach (<https://nda.nih.gov/contribute/nda-data-harmonization-approach.html>).

Procedures

1. Split REDCap summary forms into individual NDA data structures

NDA data structures are tabular representations of a single measure or assessment. Data submitted to NDA are organized by NDA data structure. REDCap may use summary forms that combine scores from many sources. REDCap output can also have all definitions in a single file. Thus, REDCap outputs may need to be split into separate NDA data structure files, by form, in order to conform to NDA data harmonization standards (<https://nda.nih.gov/contribute/harmonization-standards.html>). Submission templates for NDA data structures are required for using NDA's Validation and Upload tool. Specific transformations of this type include:

- a. Combine separate item-level data and scores of REDCap into a single NDA data structure.
- b. Combine multiple self-contained forms in REDCap into a single NDA data structure when different respondents' data are present for the same form.
 - a. Separate REDCap assessments can still be submitted, each in their own NDA templates for a user's convenience, and can be merged automatically by NDA's Validation and Upload Tool.

Example: Transformations for data structures used by the Human Connectome Project's Aging and Development studies were automated with these [Jupyter Notebooks](#) (see Step 3 for more nuanced links).

2. Merge Yes/No outputs from REDCap multichoice answers into one NDA data element.

REDCap offers multiple yes-or-no fields (0 = No & 1 = Yes) where NDA presents multichoice answers as one data element. Codes of 1 should be converted either to the text they represent or to other numeric codes. They can then be listed in a single cell of a free-text column. Where these multiple responses are bound for a single-choice NDA column, multiple rows per participant will be necessary with one value each.

3. Merge REDCap timepoint-dependent variables into one, by NDA interview_date data element.

REDCap uses separate variables for the same question at each timepoint and/or form. NDA has a single data element to track each unique timepoint. This is the element interview_date. In order to repeat an assessment question at different timepoints, transpose those data into a single column for the repeated question. Each cell will appear on its own row with a different value in the 'interview_date' column. Differentiating by visit name (e.g., screening, baseline, 3-month follow-up) or other information is also possible. Please contact the NDA Help Desk at NDAhelp@mail.nih.gov to learn more about these options.

4. Change REDCap variable names so they do not start with a number and are under 30 characters

REDCap allows any kind of field name, while NDA caps data element names at 30 characters, which cannot begin with a number, and only allow the underscore as a special character. Variable names from REDCap outputs will need to be modified if they do not comply with NDA data element name requirements.

This [REDCap-to-NDA tool](#) will evaluate REDCap data files for compliance with NDA standards.

5. Use a consistent name for each unique REDCap field. These will be assigned to NDA elements. These elements will be reused where items are repeated.

NDA can create aliases (synonyms for element names), which will allow multiple REDCap fields to correspond to a single NDA element. However, only one name for the column can appear in a data structure.

NDA does not allow references from one structure's data to another. The closest NDA can come is to import the element(s) from the other structure and repeat the data.

6. Edit REDCap field descriptions as if there were no descriptive fields.

Coding values and question stems must be repeated for each element in NDA. This may require manual work by the data manager.

7. Provide supplemental materials in your Collection's Supporting Documentation.

Visual aids present in REDCap are not translatable to NDA format. These materials can only be described or provided as Supporting Documentation on a project's Collection Page.

Supplemental resources:

- REDCap-to-NDA format converter: <https://github.com/ABCD-STUDY/redcap-to-nda>
- Conversions for REDCap data gathered using Human Connectome Project methods: https://github.com/humanconnectome/NDA_submissions
 - This includes NIH Toolbox data gathered using iPad administration.
- Converting NIH Toolbox iPad data to NDA format: <https://github.com/humanconnectome/NIHToolbox2NDA>
- Converting NDA data for use in R: <https://github.com/ABCD-STUDY/analysis-nda>