RADx-rad Data Dictionary Guide

To ensure data collected at RADx-rad are compliant with the FAIR (Findable, Accessible, Interoperable, and Reusable) principles, NIH recommends using Common Data Elements (CDEs) for data collection and data standardization. The consistent use of CDEs within the RADx program will greatly increase the interoperability and reusability of collected data, thus enabling efficient data integration and data analysis across different studies.

The RADx-rad Data Dictionary (R2D2) defines the data elements (columns) in your data files. Data elements may be common among projects, e.g., the RADx-rad Minimum CDEs, common data elements within your FOA area, or specific to your project. The Variable/Field Names in your R2D2 must match the column headers in your data files. The R2D2 may contain additional data elements that are not present in your data files, e.g., waived CDEs.

The R2D2 contains two sections:

Rows 2 – 47: RADx-rad Minimum CDEs. This section should not be changed.
Rows 48 - : FOA and/or project specific data elements

How to create the RADx-rad Data Dictionary file

1. Make a copy of the data element template file: https://www.radxrad.org/wp-content/uploads/2021/08/RADxrad_Data_Element_Template_v002-1.csv
2. Append new data elements starting in row 48, following the guidelines below.
3. Send a draft version to DCC for review and feedback.

Description of the RADx-rad Data Dictionary file

This section describes each column in the data dictionary file. The column names in the R2D2 must not be changed. Some columns in the R2D2 are specific for REDCap codebooks.

1. Column A – **Variable / Field Name** (Required)
   - Variable/Field names specify the variable name that will be used in your data files as column headers.
   - Field names
     - may contain letters, numbers, and underscores, but no spaces or special characters. Separate words by underscores, e.g., target_gene
     - cannot start with a number.
     - must be unique, and cannot be repeated within a data dictionary

2. Column B – **Section Header** (Required)

The section headers are used to group related variables, e.g., Medical History, Symptoms.

3. Column C - **Field Type** (Required)
The field type defines the type of the variables. For compatibility with REDCap codebooks, we support the categorical field types yesno, radio, dropdown, and checkbox). If you need additional field types, please contact DCC.

Field Type	**Definition**	**Notes/Examples**
text	text data	text must be quoted if it contains commas
integer	integer value	5
float	floating point number	5.0
date	YYYY-MM-DD (date in ISO 8601 format)	2021-01-01
time	hh:mm or hh:mm:ss, (24-hr format in the local time zone)	15:00
timezone	UTC±hh:mm	UTC-05:00 (New York on standard time)
zipcode	5-character ZIP code	Include leading zeros, e.g., 02101 (ZIP code in Boston)
url	reference to a web resource	https://radx-rad.org/, ftp://ftp.ncbi.nlm.nih.gov/
sequence	5'-3'-DNA, 5'-3'-RNA, or protein sequence	TAGCACTCT… This field type represents sequences of reagents, such as primers or aptamers. Sequencing data should be submitted to dbGaP.
list	list of values separated by the vertical bar | without spaces.	1|2|3 AZ|CA
category	categorical variable with a defined value set of strings	
Categorical field types from REDCap codebooks		
yesno	value must be **one** choice (integer value) from column E	1
radio	value must be **one** choice (integer value) from column E	1
dropdown	value must be **one** choice (integer values) from column E	1
checkbox	value(s) must be **one or more** choices (integer values) from column E. Multiple values must be separated by the vertical bar | without spaces	1|2|3

4. Column D – **Field Label** (Required)
A Field Label (or variable label) is a question or description of the variable/field name. Additional information can be provided in column F Field Note.

5. Column E – **Choices, Calculations OR Slider Labels** (Required for categorical field types)
The specification of choices depends on the field type.

Choices for the field types: category, list
- **Value format:** specify the response options as text.
- Example: caffeine | creatinine | sucralose | ibuprofen

Choices for REDCap field types: yesno, radio, dropdown, checkbox
- **REDCap format:** specify response options associating numerical values with labels.
- Example1: 1, Yes | 0, No
- Example2: 1, Excellent | 2, Very good | 3, Good | 4, Fair | 5, Poor

6. Column F – **Field Note** (Optional)
Field notes are used to provide additional information, e.g., special cases or links to resources.

7. Column G – **Text Validation Type OR Show Slider Number** (Optional, for REDCap only)
This field should only be used for REDCap text Field validation. Format validation types for text fields are: date, time, integer, number, zipcode.

8. Columns H & I – **Text Validation Min/Max** (Optional)
For numeric values the minimum and maximum acceptable values may be specified. The min/max values are inclusive, e.g., 0 and 100 for a variable in percent.

9. Column J – **Branching Logic** (Optional, for REDCap only)
Branching logic can be applied to a field to specify whether or not it will be displayed, depending on values in previous fields. Branching logic is typically imported from a REDCap database.

10. Column K – **Unit** (Required for numeric values that have units)
Units should be specified for numeric fields. Use SI units if feasible. For variables that represent percent values, use the word “percent” as a unit. **Do not add units to numerical data in the data files!**

11. Column L – **CDE Reference** (Optional)
A reference for the origin of the data element. The RADx-rad Interactive Mapping tool (https://www.radxrad.org/resource/interactive-mapping-interface-for-cdes) can be used to find common data elements (CDEs) from the NIH COVID and NIH ALL CDEs. If no

matching data elements can be found in the NIH CDEs, consider adding references to data elements in PhenX or other ontologies (e.g., LOINC) if they exist.