

RADx Data Hub File Organization

The RADx Data Hub is a file-based repository for data that has been collected as part of RADx Program funded research studies on COVID-19 testing. This document describes the various types of files that are hosted in the RADx Data Hub and their relationship to each other.

In the RADx Data Hub, files are grouped into two main categories: (1) Data files or image files that contain data derived from studies, and (2) Files that contain metadata that describe data files and their contents.

Data files with Tier 1 CDEs (CDEs that can be harmonized across studies and DCCs) should have an “orig” copy which reflects how the Awardee decided to code their data, as well as a “transform” copy (generated by the (C)DCC) which contains the data that has been conformed to meet the Tier 1 CDE status so that the data are already harmonized according to the [RADx Global Codebook](#). Thus, Xyz_DATA_transformcopy.csv has been derived from Xyz_DATA_origcopy.csv. Precise details of how data files are structured and the relationship between origcopy and transformcopy files are described in [RADx Data File Naming Conventions and Content Rules for Data Submitters](#).

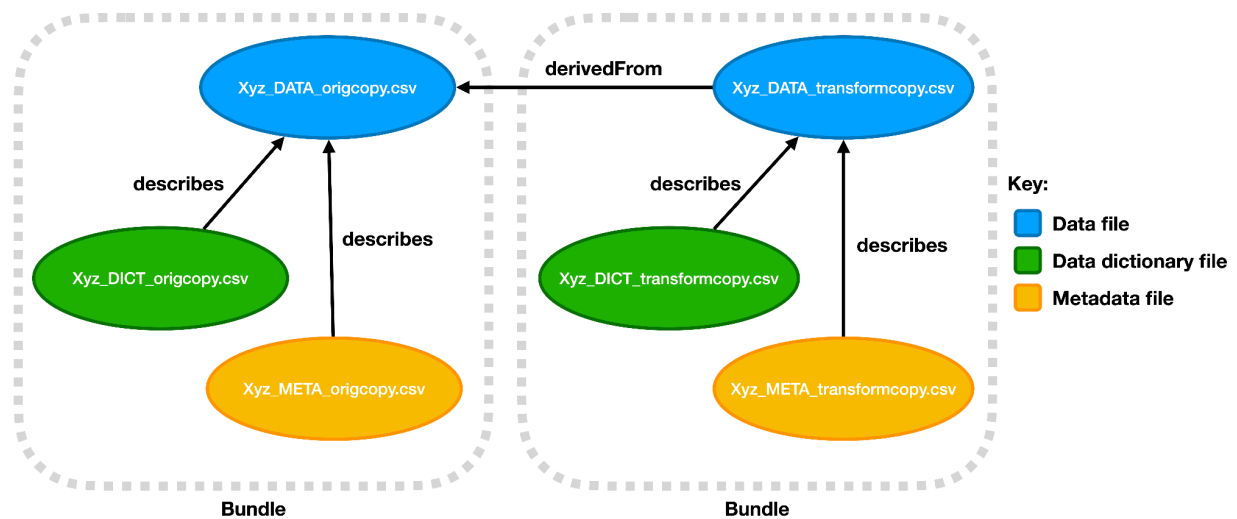
Bundles

File bundles group together a data file (DATA), its data dictionary file (DICT), and its general metadata file (META). Note that both the META and DICT files, described in more detail below, are files containing metadata. Complete bundles are required for submission into the RADx Data Hub. Files are most easily linked together by file naming conventions. For a study that has produced some data under the title “Xyz”, the RADx Data Hub contains at least the first three files as shown in the table below. If the study involves human participants and provides RADx core data elements, these will have been harmonized in accordance with the RADx Global Code Book. In this case a bundle will also contain transformcopy files as shown in the last three rows of the table below.

File Name	Category	Related to file
Xyz_DATA_origcopy.csv	Data	–
Xyz_META_origcopy.json	Metadata	Xyz_DATA_origcopy.csv
Xyz_DICT_origcopy.csv	Metadata	Xyz_DATA_origcopy.csv
Xyz_DATA_transformcopy.csv	Data	Xyz_DATA_origcopy.csv
Xyz_META_transformcopy.json	Metadata	Xyz_DATA_transformcopy.csv

The above files and their relationships can be drawn out using a node-arc diagram as shown below. Bundles are depicted by the light gray dashed lines. There is a bundle for the _origcopy files and a bundle for the _transformcopy files.

File Relationships



The different types of files are described in more detail below.

DATA files

Study data are typically uploaded as Comma Separated Values (CSV) files that contain tabular data and are denoted with “_DATA” in the file name. For more information see [RADx Data Files and Conventions](#).

DICT files

Data dictionary files that describe the columns in a DATA file that contains tabular data are typically denoted with “_DICT” in the file name. If a DATA file does not contain tabular data, then it will not have an accompanying DICT file. Data dictionaries are Comma Separated Values (CSV) files that follow the [RADx Data Dictionary Specification](#). To view an example data dictionary, please see the [RADx Data Hub Global Code Book Data Dictionary](#).

META files

Tabular data files will have a second accompanying metadata file formatted as a JSON file and containing JSON-LD structured descriptions of DATA file attributes and contents. These files are identified with “_META” in the file name. The format and structure of META files is described in detail in the [RADx Metadata Specification](#).

Study Documents

The RADx Data Hub can also support non-structured supplemental data files that live outside of specific bundles. These files provide additional details and context to researchers to allow the appropriate use of study data. Study documents may include, but are not limited to, study protocols, data collection forms, or README files provided by study staff. These types of files are not shown in the above diagram.