

Psych-DS Introduction

Psych-DS Standard: https://psychds-docs.readthedocs.io/en/latest/reference/rules_and_conventions/

psychds package: <https://github.com/psych-ds/psychds-r>
psychds documentation: <https://psych-ds.github.io/psychds-r/>

CEDAR Psych-DS form: <https://psych-ds.github.io/cedar-wizard-psychds/>

Psych-DS Overview

Psych-DS is a community data standard inspired by the Brain Imaging Data Set (BIDS). A data standard is an agreed upon set of rules for the organization and documentation of data.

In formalized data standards like Psych-DS, it is typical to use a schema to make these rules machine-readable and enable validation of the standard. The primary rules of the Psych-DS package include

1. Structure of your directory:
 - must have a data/ directory within your project directory
 - raw data is stored in data/raw/ and not considered in validation
 2. Format of your data:
 - in your data/ directory, must have a data file saved as a .csv file
 3. Metadata documentation:
 - dataset_description.json required with minimum metadata fields:
 - i. “name” – dataset name
 - ii. “description” – dataset description
 - iii. “variableMeasured” – list of all column headers/names in your data/dataset.csv file
-
- Recommended additional metadata fields to make a JSON-LD file:

- i. Make it a JSON-LD by including namespace and type



- ii. Author(s)

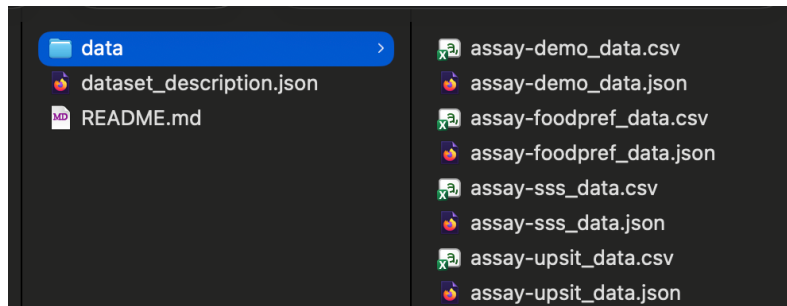
```
"author" : [  
  {  
    "@type" : "Person",  
    "name" : "Alaina L Pearce"  
  },  
  {  
    "@type" : "Person",  
    "name" : "Author name 2"  
  },  
]
```

- iii. Add a variable dictionary rather than just a list of names

```
" variableMeasured " : [  
  {  
    "name" : "id",  
    "description" : "participant id",  
    "required" : "true",  
    "@type" : "PropertyValue"  
  },  
  {  
    "name" : "age",  
    "description" : "participant age",  
    "unitText" : "years",  
    "minValue" : "10",  
    "maxValue" : "16",  
    "@type" : "PropertyValue"  
  },  
  {  
    "name" : "sex",  
    "description" : "participant sex",  
    "valueReference" : [  
      {  
        "value" : "0",  
        "label" : "male"  
      },  
      {  
        "value" : "b1",  
        "label" : "female"  
      }  
    ],  
    "@type" : "PropertyValue"  
  },  
]
```

Psych-DS for Complicated Datasets

While Psych-DS can be a very simple way to document smaller, less complex datasets, it can also be leveraged for more complicated datasets by the use of sidecars. Sidecars are additional JSON metadata files for datasets that involve multiple data files or multiple different types of data.



1. Metadata documentation:

- dataset_description.json required (with suggested additional metadata):

```
"@context" : "http://schema.org/",
"@type" : "Dataset"
"schemaVersion" : "Psych-DS 0.1.0",
"author" : [
  {
    "@type" : "Person",
    "name" : "Alaina L Pearce"
  }
],
"temporalCoverage" : "2022-2026"
```

- (recommended) include list of data with description

```
"hasPart": [
  {
    "@type": "Dataset",
    "name": "Data 1",
    "description": "data 1 description",
    "url": "data/data1.csv",
    "encodingFormat": "text/csv"
  },
  {
    "@type": "Dataset",
    "name": "Data 1",
    "description": "data 2 description",
    "url": "data/data2.csv",
    "encodingFormat": "text/csv"
  }
]
```

- Sidecars – each data file needs their own JSON file that matches the name of the data file with all variables listed

Using psychds

1. Identify a dataset
2. Open RStudio (install instructions: <https://posit.co/download/rstudio-desktop>)
3. Install packages

```
install.packages("remotes")
```

```
remotes::install_github("psych-ds/psychds-r", dependencies = TRUE)
```

4. Open Shiny app

```
library(psychds)
```

```
run_psych_ds_app()
```

Tools in This App



Create Dataset

Choose some existing data on your device. A step-by-step interface will help you build a project folder with the same data in Psych-DS format, and then save the whole directory to your machine. This will not change your original data in any way unless you choose to overwrite the directory your data came from.



Validate Dataset

Check if a dataset complies with the Psych-DS standard. You can check the dataset you just made, or use this on a directory that you or someone else edited by hand.



Update Dictionary

Psych-DS uses a small text file to store your metadata (information about your dataset). This text file includes an entry for every variable (column) in your data. Use this page to automatically edit your metadata with variable definitions and other information, or to export a nicely formatted PDF of your data dictionary.



Dataset Explorer

Because Psych-DS datasets are all structured in the same basic way, we can build tools that work on any Psych-DS dataset. Use this page to browse your data and see filters, summaries about missing data, and more.



Upload to OSF

Upload a Psych-DS dataset to the Open Science Framework (OSF), auto-filling all the information you've already provided about this dataset (descriptions, authors, etc.) Beta

5. Select Create Dataset

- Step 1 – set up directories
 - i. Name your project
 - ii. Select which file contains your data

- iii. Select the data file(s) to use
- iv. Decide on the types of subfolders you would like to use

Name Your Project Directory

The goal of Psych-DS is to standardize how you store data within a scientific project. This tool will build a new project directory to store your data, with additional folders (analysis, materials, etc.) if you like.

Select Data Directory

Choose the folder on your computer that contains all the data files you want to include in your new project directory. It's okay if that folder also contains other things; you'll select the specific data files below.

Select Data Files

Select the CSV or TSV files that you want to include in your Psych-DS data folder. If your data are not yet in CSV or TSV format, you'll need to start by [converting them](#).

Click files to select them individually, or click directory names to select/deselect all files in that directory.

☒ Select All
 ☐ Deselect All

Please select a project directory first

Optional Subfolders

This tool can create additional empty directories inside your project folder for you to move your other materials into.

☒ analysis/
 ☒ materials/
 ☐ results/

Add a custom subdirectory:

- Step 2 - metadata
 - i. Review variable names – if need to change/update, will need to do it in the file and start over
 - ii. Add dataset name and description
 - iii. Add authors

Step 2: Dataset Metadata

Every Psych-DS project has a text file called `dataset_description.json` with information (metadata) about the dataset. This file is part of what makes Psych-DS work - it sits in your project directory and is used to confirm whether your specific dataset matches the standard.

Right now, we'll make you a basic version of this file that you can update later with more information. You can do this by either editing the text file by hand or using the "Update Dictionary" tool.

Review Detected Variables

Psych-DS uses your column headers to create the list of variables in your dataset. Variable names used in more than one file are assumed to have identical definitions. Make sure the variable names you see make sense for your data files!

Search:

variable	present_in
cond	<input type="text" value="whole_vs_parts.csv"/>
h2og	<input type="text" value="whole_vs_parts.csv"/>
id	<input type="text" value="whole_vs_parts.csv"/>
sandg	<input type="text" value="whole_vs_parts.csv"/>
sandkcal	<input type="text" value="whole_vs_parts.csv"/>

Showing 1 to 5 of 6 entries Previous 1 2 Next

Dataset Information

Name *

Description *

Briefly describe your dataset

Author Information

Name	ORCID ID
No authors added yet. Click 'Add New Author' below.	

- Step 3 – standardize filenames
 - i. Check file names
 - ii. If not valid, choose key to use in key-value pair naming and rename file
 - iii. Select where to save data and review structure

Step 3: Standardize Filenames

Psych-DS has specific naming conventions that your data files need to follow. This naming system is going to make you explain what each piece of a filename means: you can't just say "347B". Instead you have to use keywords to describe what that refers to. Is that participant 347B? Session 347B? Or even participant 347, session B?

We encourage you to use keywords from the suggested list below, but you can add your own if needed.

Already renamed your files?
 If your filenames already follow the Psych-DS pattern (e.g., `subject-01_task-memory_data.csv`), click to validate them automatically.

Validate Existing Filenames

1. Select Files

Click on one file to create the name it will have in your Psych-DS folder, or use the check boxes to choose a batch of files to rename at the same time.

Select All

Deselect All

Original Filename	New Filename
<input type="checkbox"/> whole_vs_parts.csv	← Configure keywords

Files: 0 of 1 selected
Configuring: whole_vs_parts.csv

Auto-Name from Data
 Fill keyword values from columns in your data

 No constant-value columns detected in selected files.

2. Choose Keywords

Select keywords to include in your filename. Drag to reorder.

subject

session

study

task

condition

stimulus

trial

description

Add Custom

3. Keyword Values

Select keywords above to configure values

6. Validate Dataset

- Select dataset you just created and validate

7. Update Dictionary

- Load dataset
- Select each variable and update:
 - i. Description
 - ii. Data type
 1. Factor information if needed

iii. Measurement information

- Update list of missing values code to reflect your data

8. Explore dataset

- Load dataset
- Select variable to get summary/descriptive information

Search

Filter Files by Keyword/Value

Add

Filter Rows by Column/Value

Add

Active Filters

Keyword Filters

No keyword filters active

Column Filters

No column filters active

Variable Statistics

sandg

Total 104	Unique 104
Missing 0 (0%)	Valid 104

Numeric Statistics

Mean 266.981	Median 256.7
Std Dev 108.847	Range 86.6 to 537.5