

Getting Started Creating Data Dictionaries: How to Create a Shareable Dataset

Erin M. Buchanan^{12abc}

Sarah E. Crain^{1abc}

Arielle Cunningham^{1abc}

Hannah R. Johnson^{1abc}

Hannah Stash^{1abc}

Marietta Papadatou-Pastou^{3c}

Peder Mortvedt Isager^{4c}

Rickard Carlsson^{5c}

Balazs Aczel^{6c}

¹Missouri State University

²Harrisburg University of Science and Technology

³National and Kapodistrian University of Athens

⁴Eindhoven University of Technology

⁵Linnaeus University

⁶Eötvös Loránd University

Corresponding author: Erin M. Buchanan (ebuchanan@harrisburgu.edu)

Disclosures:

Author contributions: ^aDrafted the manuscript, ^bContributed to the application development, ^cCommented or edited on applications or manuscript

Conflicts of Interest: The authors declare that there were no conflicts of interest with respect to the authorship or the publication of this article.

Acknowledgments: The authors would like to thank the Psych-DS team for their comments and contributions to the manuscript and applications.

Prior versions: The pre-print link is <https://osf.io/vd4y3/>.

Data, materials, and online resources: The data and materials can be found at

<https://osf.io/3y2ex/>. The code for the applications can be found at

<https://github.com/doomlab/data-dictionary/>.

Abstract

As researchers begin to embrace open and transparent data sharing, they can be left wondering how to properly detail their data. Simple sharing on a repository website is an easy first step, however, without proper documentation, these files can be completely missed by other researchers and indexing search engines. This tutorial demonstrates three different applications that allow a researcher to create a data dictionary for a dataset. Data dictionaries provide a wealth of information about the variables, data collection, and other important facets of the dataset. This information, or metadata, provides key insights into how the data might be further used in research and allows for search engine indexing to reach a broader audience of interested parties.

Getting Started Creating Data Dictionaries: How to Create A Shareable Dataset

A data dictionary is a supplementary document that details the information provided in a dataset, often the meaning of included variables, creation, format, and usage of the data (McDaniel & International Business Machines Corporation, 1994). This information is usually denoted as metadata. This tutorial will demonstrate three simple options for creating data dictionaries to describe research data in psychology (see psych-DS for current efforts; Kline, 2018). These applications allow the researcher to enter their metadata with the goal of sharing the files on a platform for others to read. In general, metadata should describe the information provided in the dataset, and websites like schema.org provide structure for what can and should be included in a data dictionary.

Open datasets are beneficial for both individual researchers and the scientific community as a whole: articles with open datasets reach more researchers, allowing them to convey their findings to a wider audience, as evident by their rate of citations (McKiernan et al., 2016). Open data allow scientists to develop new hypotheses (e.g., Vadillo, Gold, & Osman, 2018), investigate multiple analysis perspectives on different datasets (e.g., Simonsohn, Simmons, & Nelson, 2015), and, importantly, identify and correct errors that would otherwise create noise in the literature (Piwowar & Vision, 2013). FAIR guidelines indicate that data should be Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016). One of the most common researcher concerns for sharing data is not having set standards for making data public (Hardwicke et al., 2018; Houtkoop et al., 2018). For those wishing to use the open data, one concern is that the shared data are unusable without a data dictionary to translate the data.

Further, open data may not be findable without a corresponding, machine-readable, data dictionary containing the searchable metadata.

This tutorial aims to help researchers by demonstrating how to create metadata for their datasets. Table 1 indicates the practical benefits to each of the applications described below. We have also included video tutorials as supplementary material, and the links are provided in Table 1 (also online at <https://osf.io/3y2ex/>). The allowed inputs into the applications cover a wide range of data formats including SPSS/SAS data files, comma separated files (CSV), text files, and Excel file formats. The output from these applications is HTML (i.e., webpage), CSV, and JavaScript Object Notation (JSON). JSON is a machine readable format, and thus searchable, which are encouraged for sharing, especially by the new Google Dataset Search (Noy, 2018). Once the data dictionary is created, these files can be shared alongside the dataset on a web repository in the same folder (see Rouder, 2016 for a tutorial)¹. In the case of multiple datasets and dictionaries, separate subfolders or naming cues should be included to ensure that researchers can map the dataset to the appropriate dictionary.

Table 1.

Information about the Benefits for each Application

	Codebook	DataSchema	DD Creator
Citation	Arslan (2018)	Inspired by Data Spice (Boettiger et al., 2018)	DeBruine, Buchanan, & Mohr (2018)
Input	CSV, SPSS, Stata, RDS	CSV, Text, Excel, SPSS, SAS	CSV, Text, Excel, SPSS, SAS
Output	HTML report from Markdown	CSV files of metadata, JSON, and HTML report	CSV files of metadata, JSON, Rdata
Benefits	Easiest to use	Follows schema.org	Follows schema.org

¹ We do not specify where to share the data and metadata as there are many options that are appropriate. However, a tutorial for the Open Science Framework can be found here: <https://youtu.be/Uyw42vpXCPk>.

	Quick metadata generation Generates a summary for each variable in a readable format	Metadata entry is medium	Specifies a separate section for category labels Rdata output More detailed descriptions, depending on the data
Video	youtu.be/GF-mPLnPD1g	youtu.be/GRdMjrDurOY	youtu.be/7gZIz1tvQ_Y

Data Files

In this tutorial, we demonstrate examples of how to process a dataset and create different types of metadata output. The supplementary video tutorials describe each input space and provide examples of possible descriptions. The *mtcars* data example is from *R*'s *datasets* package and contains data from car road tests for $n = 32$ cars on ten variables (Henderson & Velleman, 1981). The requirements for data in these tutorials is that they: 1) be in a file format that is readable by one of the demonstrated apps as shown in Table 1, 2) include participants², and 3) include variables. Data may be arranged according to tidy data principles; however, these apps do not specify the layout of the data, only that it may be labeled and described as shown (Wickham, 2014).

Codebook Tutorial

Codebook (Arslan, 2018) is an *R* package with a corresponding website that allows researchers to create reports of their data, including reliabilities and summaries of items (histograms, descriptive statistics). The embedded metadata (such as item labels) is automatically combined into that report. Of the three available options, codebook is the quickest and easiest to implement; however, users who are accustomed to Graphical User Interfaces in their statistical software (e.g., SPSS users) might have trouble editing the automatically produced output if they

² Participants are defined as a person/animal from a study or other types of studied data, such as stimuli, words, etc.

wish to add more information. The tutorial below covers the web app version of Codebook, and Arslan (2018) provides more information about using Codebook in *R*.

Step 1: Go to the Codebook website: <https://opencpu.formr.org/ocpu/library/codebook/www/>.

Codebook generator

SPSS/Stata/RDS File:
 No file chosen

```
1  ---
2  title: "Codebook"
3  output:
4    html_document:|
5      toc: true
6      toc_depth: 4
7      toc_float: true
8      code_folding: 'hide'
9      self_contained: true
10   pdf_document:
11     toc: yes
12     toc_depth: 4
13     latex_engine: xelatex
14   ---
15
16   ```{r setup}
17   knitr::opts_chunk$set(
18     warning = TRUE, # show warnings during codebook generation
19     message = TRUE, # show messages during codebook generation
20     error = TRUE, # do not interrupt codebook generation in case of errors
21     # usually better for debugging
22     echo = TRUE # show R code
23   )
24   ggplot2::theme_set(ggplot2::theme_bw())
25   pander::panderOptions("table.split.table", Inf)
26   ```
27
28  We collected the following data.
```


Step 2: Click the **Browse...** button at the top of the page. Select the data file you wish to upload, then click **Open**. You may upload CSV, SAV (SPSS), DTA (Stata), or RDS files.

Codebook generator

SPSS/Stata/RDS File:

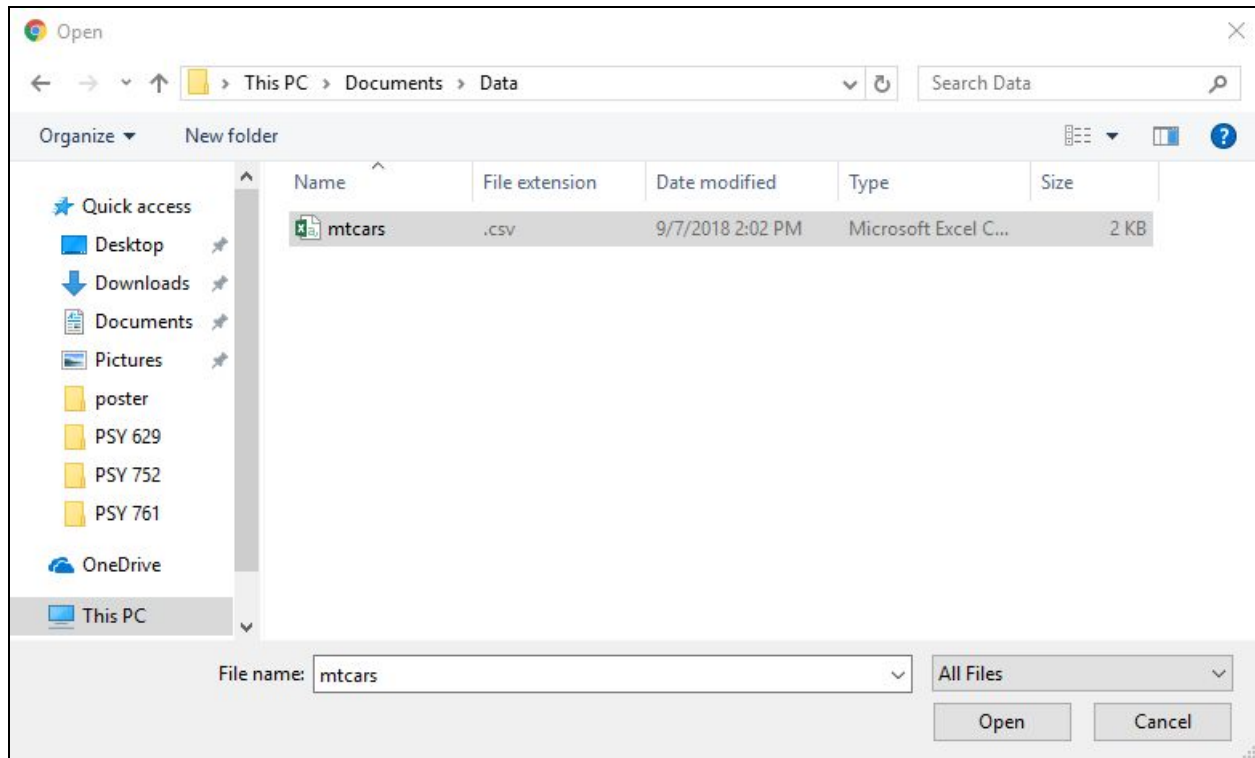
Choose File

No file chosen

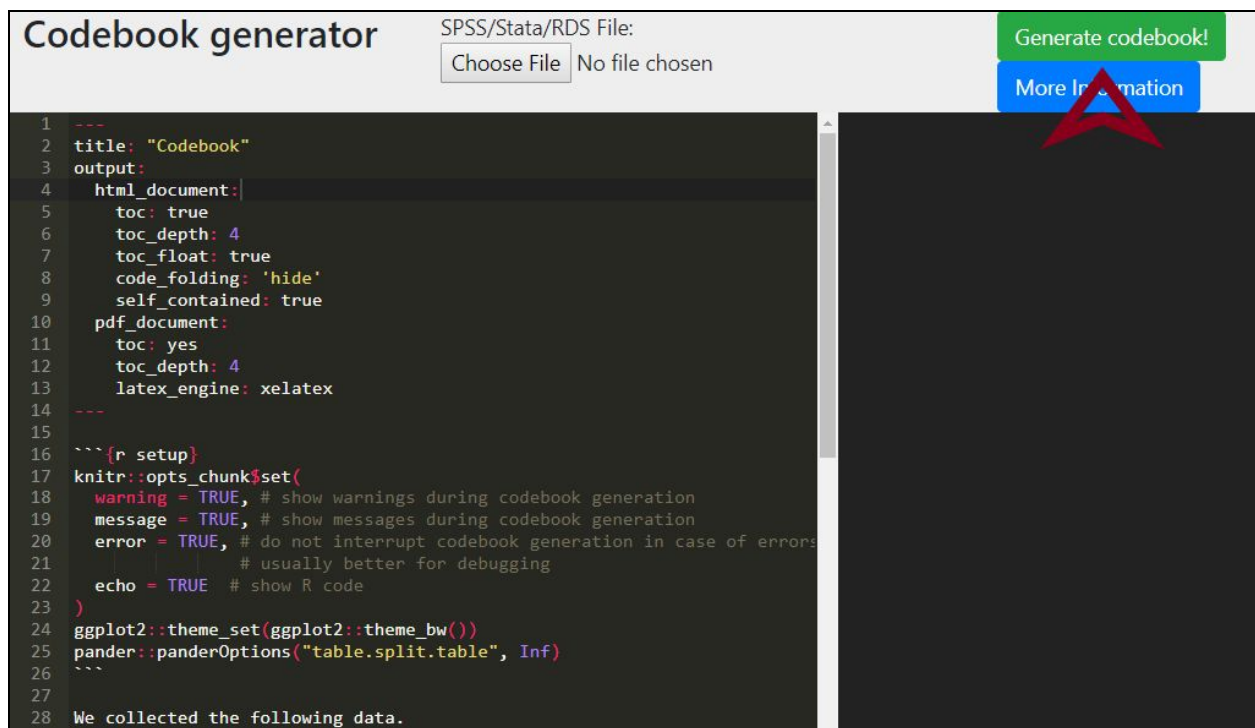
Generate codebook!

More Information

```
1 ---
2 title: "Codebook"
3 output:
4   html_document:
5     toc: true
6     toc_depth: 4
7     toc_float: true
8     code_folding: 'hide'
9     self_contained: true
10  pdf_document:
11    toc: yes
12    toc_depth: 4
13    latex_engine: xelatex
14 ---
15
16 ```{r setup}
17 knitr::opts_chunk$set(
18   warning = TRUE, # show warnings during codebook generation
19   message = TRUE, # show messages during codebook generation
20   error = TRUE, # do not interrupt codebook generation in case of errors
21   # usually better for debugging
22   echo = TRUE # show R code
23 )
24 ggplot2::theme_set(ggplot2::theme_bw())
25 pandoc::pandocOptions("table.split.table", Inf)
26 ```
27
28 We collected the following data.
```



Step 3: Click **Generate codebook!**



Step 4: Make changes to the created document by editing the code on the left side of the screen.

You can safely edit the title, the text that says, “We collected the following data.”, or “My

Awesome Dataset” for the name of the dataset. Those more familiar with R Markdown may edit elsewhere. To save changes, click **Generate codebook!** again.

```

1  ---
2  title: "Codebook"
3  output:
4    html_document:
5      toc: true
6      toc_depth: 4
7      toc_float: true
8      code_folding: 'hide'
9      self_contained: true
10   pdf_document:
11     toc: yes
12     toc_depth: 4
13     latex_engine: xelatex
14   ---
15
16   ```{r setup}
17   knitr::opts_chunk$set(
18     warning = TRUE, # show warnings during codebook generation
19     message = TRUE, # show messages during codebook generation
20     error = TRUE, # do not interrupt codebook generation in case of errors
21     # usually better for debugging
22     echo = TRUE # show R code
23   )
24   ggplot2::theme_set(ggplot2::theme_bw())
25   pandoc::pandocOptions("table.split.table", Inf)
26   ```
27
28   We collected the following data.
29

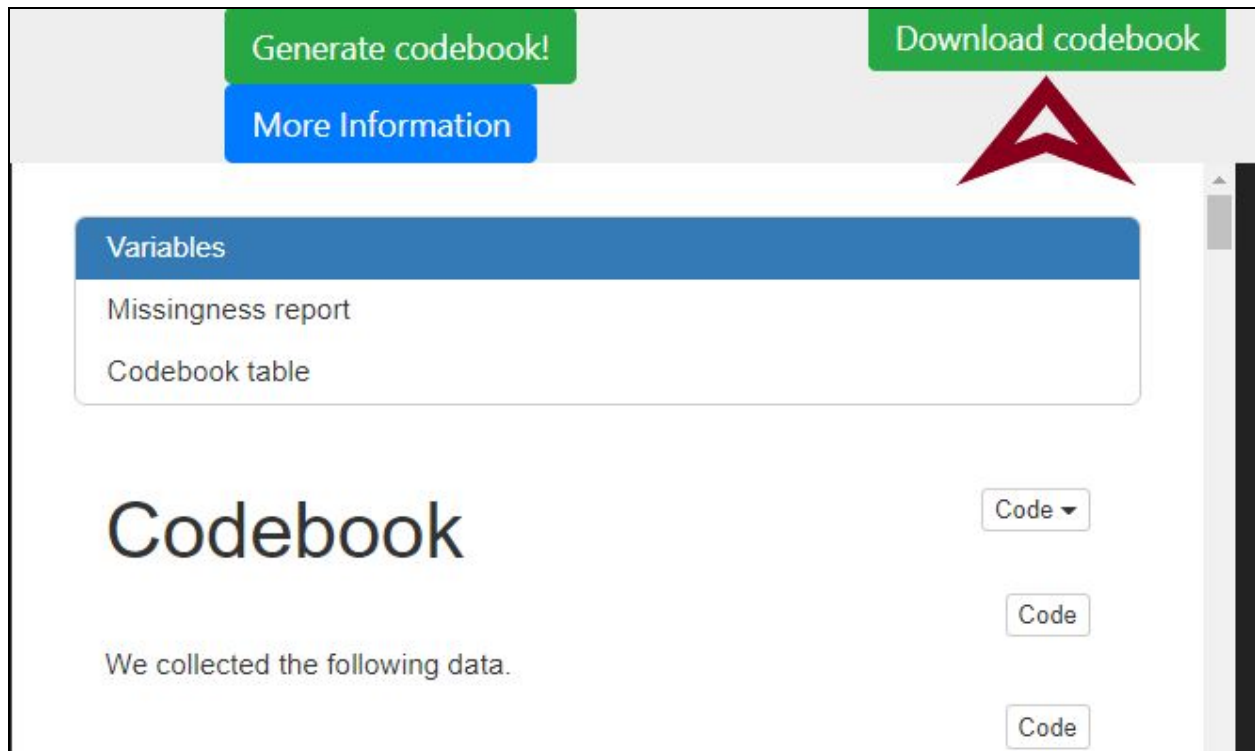
```

```

46  codebook_data = detect_scales(codebook_data)
47
48
49  # Does your dataset have a name that is not reflected in the file name?
50  metadata(codebook_data)$name <- "My Awesome Dataset"
51
52  codebook(codebook_data)
53  ```
54

```

Step 5: When ready, click **Download codebook** to save the file to your computer.



The screenshot shows a web interface for generating and downloading a codebook. At the top, there are two green buttons: "Generate codebook!" and "Download codebook". Below the "Generate codebook!" button is a blue button labeled "More Information". A red arrow points to the "Download codebook" button. On the left side, there is a sidebar with a blue header "Variables" and two items: "Missingness report" and "Codebook table". The main content area has a large heading "Codebook" and the text "We collected the following data." Below this text are three buttons, each labeled "Code". The first button has a dropdown arrow, while the other two are plain.

Generate codebook!

Download codebook

More Information

Variables

Missingness report

Codebook table

Codebook

We collected the following data.

Code ▼

Code

Code

DataSchema Tutorial³

This application allows users to create an HTML report, a JSON file formatted following guidelines for datasets from schema.org, and .csv files of their metadata. In this application, descriptions of the dataset properties (e.g., authors, collection dates) and column information should be entered to complete the metadata files.

Step 1: Go to <https://doomlab.shinyapps.io/dataschema/> to access DataSchema.

Step 2: Click **Browse...** and select the data file you wish to upload. You may upload a TXT, CSV, XLS(X), SAV, or SAS data file.

DataSchema

- 1. Upload Data
- 2. Edit Attributes
- 3. Edit Access
- 4. Edit Bibliography
- 5. Edit Creators
- 6. Create json
- 7. View Report

Upload Data

This Shiny app was created to help researchers create a data dictionary that supplies the reader with information about a dataset that the researcher wishes to share. You will want to first upload your data, and then follow the steps on the left side panel in order. At the end of following the steps, you will have .csv files of the different schema.org requirements for datasets: attributes, access, bibliography, and creators. In addition, the app generates a .json file of this information to create a machine readable document for the dataset. Last, you can generate an HTML report of the information you entered to share with others for maximum reach to all audiences.

.TXT/CSV/XLS(X) Data File

Browse... mtcars.csv

Upload complete

☒ My data file has a header.

Here's a preview of your uploaded data:

Show 10 entries Search:

³ DataSchema was created by forking (copying) from the GitHub repository for Data Spice (Boettiger et al., 2018). The original R code was then rewritten completely as a single Shiny application (Chang, Cheng, Allaire, Xie, & McPherson, 2018).

mpg	cyl	disp	hp	drat	wt	qsec	vs
21	6	160	110	3.9	2.62	16.46	0
21	6	160	110	3.9	2.875	17.02	0
22.8	4	108	93	3.85	2.32	18.61	1
21.4	6	258	110	3.08	3.215	19.44	1
18.7	8	360	175	3.15	3.44	17.02	0
18.1	6	225	105	2.76	3.46	20.22	1
14.3	8	360	245	3.21	3.57	15.84	0
24.4	4	146.7	62	3.69	3.19	20	1
22.8	4	140.8	95	3.92	3.15	22.9	1
19.2	6	167.6	123	3.92	3.44	18.3	1

Showing 1 to 10 of 32 entries

Previous

1

2

3

4

Next

Step 3: Click the **Edit Attributes** page on the left to navigate to a page where you can add descriptions to the variables provided from the data file. The data uploaded will be populated into the information table on this page.

DataSchema

1. Upload Data

2. Edit Attributes

3. Edit Access

4. Edit Bibliography

5. Edit Creators

6. Create json

7. View Report

Create the Attributes Table

The attributes table describes each of the available variables provided in data. In this table, you will create a description of each column of data, which is usually a different variable from the study. You can also add information about the scale of the data and value labels for the points in the data. Click Save Changes when you are done inputting information before you move on to the next tab.

Right-click on the table to delete/insert rows. Double-click on a cell to edit. Changes are saved while on this page.

	fileName	variableName	description	unitText
1	mtcars	mpg	Miles per gallon	
2	mtcars	cyl	Number of cylinders	
3	mtcars	disp		
4	mtcars	hp		
5	mtcars	drat		

Edit the table by selecting a cell and typing directly into it. Make sure to click **Save Changes** to save the table before moving onto the next step.

Right-click on the table to delete/insert rows. Double-click on a cell to edit. Changes are saved while on this page.

	fileName	variableName	description	unitText
1	mtcars	mpg		
2	mtcars	cyl		
3	mtcars	disp		
4	mtcars	hp		
5	mtcars	drat		
6	mtcars	wt		
7	mtcars	qsec		
8	mtcars	vs		
9	mtcars	am		
10	mtcars	gear		
11	mtcars	carb		

Hit the Save Changes button to save this file.

Attributes Metadata:

Save Changes

Below the table is a box with descriptions of what to enter for each column. Each table entry page of the app contains a similar box of descriptions.

Hit the Save Changes button to save this file.

Attributes Metadata:

fileName = The name of the input data file(s). Do Not Change.

variableName = The name of the measured variable. Do Not Change.

description = A written description of what that measured variable is.

unitText = The units the variable was measured in, as well as any variable labels for data to help the reader understand the measurement of the data. For example, you might include the label values for a Likert-type scale that participants used to rate the data.



Save Changes

Step 4: Click the **Edit Access** page on the left to navigate to a page where you can include the name, type, and location of the data file. Edit the table by directly typing in the desired box, and then click **Save Changes** before moving on to the next step.

DataSchema

1. Upload Data

2. Edit Attributes

3. Edit Access

4. Edit Bibliography

5. Edit Creators

6. Create json

7. View Report

Create the Access Table

The access table helps the reader know the name, type, and location of the data file. Follow the key below in order to know what to include in each column. Click Save Changes when you are done inputting information before you move on to the next tab.

Right-click on the table to delete/insert rows. Double-click on a cell to edit. Changes are saved while on this page.

	fileName	name	contentUrl	fileFormat
1	mtcars	mtcars		csv

Hit the Save Changes button to save this file.

Save Changes

Access Metadata:
fileName = the filename of the input data

Step 5: Click the **Edit Bibliography** page on the left to navigate to a page where you can add information about the data collection and data citation. Edit the table by clicking directly on the box, and make sure to click **Save Changes** before moving on to the next step.

DataSchema

- 1. Upload Data
- 2. Edit Attributes
- 3. Edit Access
- 4. Edit Bibliography**
- 5. Edit Creators
- 6. Create json
- 7. View Report

Create the Bibliography Table

The bibliography table gives the reader information about how to cite and use your data, as well as information about the data collection. Click Save Changes when you are done inputting information before you move on to the next tab.

Right-click on the table to delete/insert rows. Double-click on a cell to edit. Changes are saved while on this page.

	title	description	datePublished	citation	keywords	lic
1	mtcars	Motor Trends data	08-24-2018		car, vehicle, motor	

Hit the Save Changes button to save this file.

Save Changes

Bibliographic metadata:

title = Title of the dataset(s) described

Step 6: Click the **Edit Creators** page on the left to navigate to a page where you can include citation information about the researchers who were involved in data collection and creation.

Edit the table by typing directly in it, and click **Save Changes** before moving on to the next step.

DataSchema

≡

1. Upload Data

2. Edit Attributes

3. Edit Access

4. Edit Bibliography

5. Edit Creators

6. Create json

7. View Report

Create the Creators Table

The creators table includes a list of researchers who participated in data collection and creation for citation purposes. Click Save Changes when you are done inputting information before you move on to the next tab.

Right-click on the table to delete/insert rows. Double-click on a cell to edit. Changes are saved while on this page.

	id	givenName	familyName	affiliation	email
1	your ORC ID	Erin	Buchanan	Missouri State University	

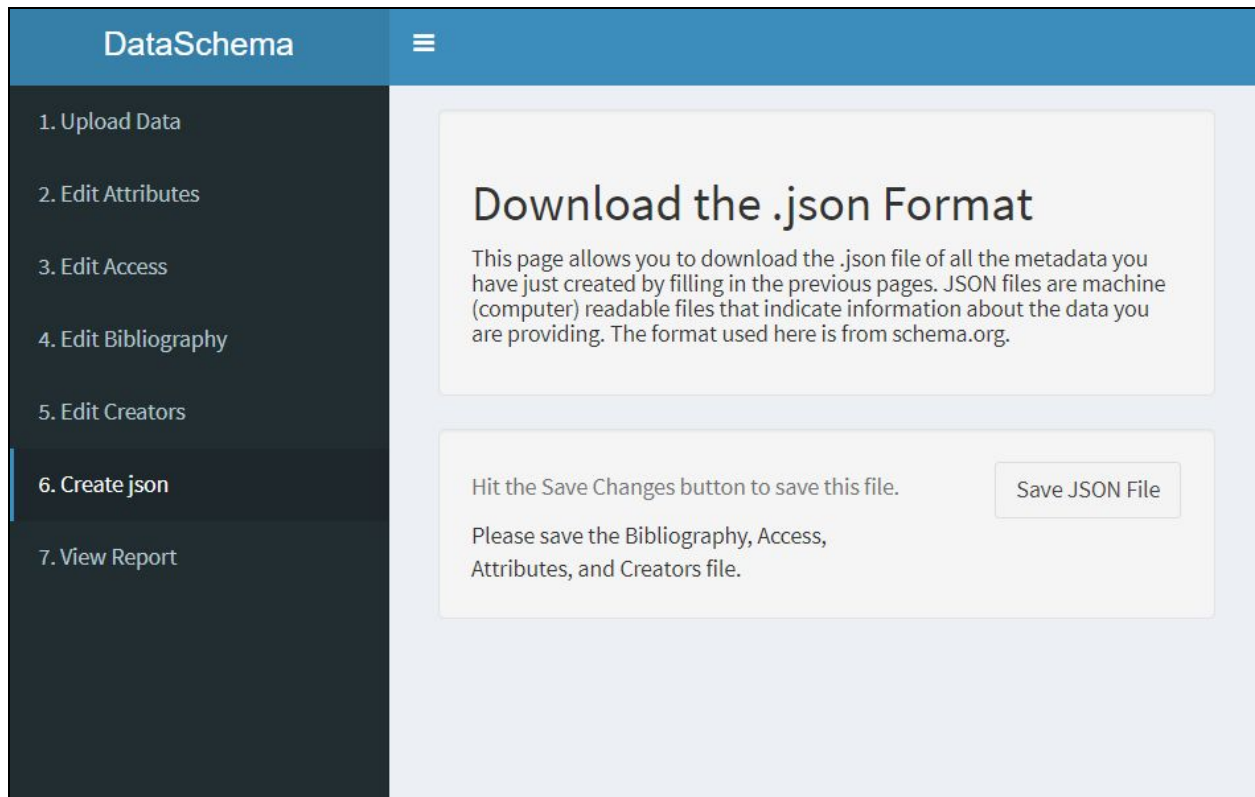
Hit the Save Changes button to save this file.

Save Changes

Creators Metadata:

id = Your ORCID ID, which you can create at

Step 7: Click the **Create json** page on the left to navigate to a page where you can download the .json file of the metadata created by editing the tables from previous pages. Click **Save JSON File** to save the metadata as a .json file.



Step 8: Click the **View Report** page on the left to view the metadata as an HTML report. Click **Save Report HTML** to save the report as a .html file. Open the file to view the HTML report of the metadata.

Dataset Information								
Title	Description	Date Published	Citation	Keywords	License	Funder	Geographic Information	Start Date
mtcars	Motor Trends data	08-24-2018	NA	car, vehicle, motor	NA	Missouri State University	Springfield, MO	03/27/2018

Accessing the Data			
File Name	Name of Data	URL	File Format
mtcars	mtcars	http://mydata.com/data	csv

Authors				
ORC-Id	First Name	Last Name	Affiliation	Email
12345	Erin	Buchanan	Missouri State University	email@missouristate.edu

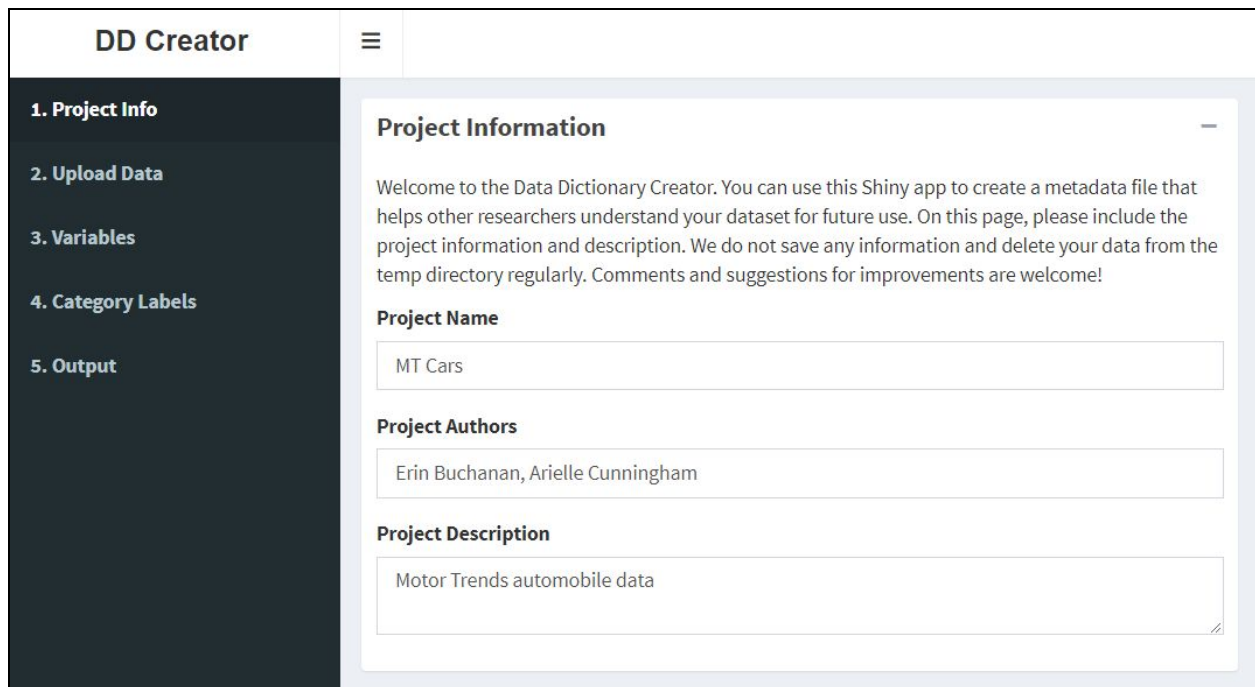
Data Attributes		
Variable Name	Description	Units of Measure

Data Dictionary Creator

DD Creator allows a user to enter metadata for each column provided in the dataset, while automatically providing a starting point for the number of unique values, missing values, variable type (i.e., character, numeric), and minimum/maximum values. A description of each column can be added, along with information about the levels/groups in the data and synonyms for the variables. On a separate page, category labels can be provided for both character and numeric data (i.e., Likert-type scales that include labeled numbers). Users can then download CSV files of the metadata, a JSON formatted metadata file, and an Rdata file that includes the dataset and descriptive information integrated together.

Step 1: Go to <https://doomlab.shinyapps.io/ddcreator/> to access DD Creator.

Step 2: On the first page (**Project Info**), type in the appropriate information under the **Project Name**, **Project Authors**, and **Project Description** fields.



The screenshot shows the 'DD Creator' application interface. On the left is a dark sidebar with a menu containing five items: '1. Project Info', '2. Upload Data', '3. Variables', '4. Category Labels', and '5. Output'. The main content area is titled 'Project Information' and contains a welcome message: 'Welcome to the Data Dictionary Creator. You can use this Shiny app to create a metadata file that helps other researchers understand your dataset for future use. On this page, please include the project information and description. We do not save any information and delete your data from the temp directory regularly. Comments and suggestions for improvements are welcome!'. Below the message are three input fields: 'Project Name' with the text 'MT Cars', 'Project Authors' with the text 'Erin Buchanan, Arielle Cunningham', and 'Project Description' with the text 'Motor Trends automobile data'. Each field has a small icon in the bottom right corner, likely for clearing the input.

Step 3. Click the **Upload Data** tab on the left side of the page. Click **Browse** and select the data file you would like to upload. If your data file has a header, check the **Data file has a header** box.

DD Creator

- 1. Project Info
- 2. Upload Data**
- 3. Variables
- 4. Category Labels
- 5. Output

Upload Data

On this page, you can upload the data file to begin making your data dictionary. Currently, this app supports text files that are comma separated, Excel files, SPSS data files, and SAS data files. After you upload your file, you can begin editing the metadata on the next pages under items 3 and 4 on the left hand menu. The data file you uploaded will be displayed below, so you can verify it was read properly.

TXT/CSV/XLS(X)/SAV/SAS Data File

Browse... mtcars.csv

Upload complete

☒ Data file has a header

A table of your data will display below. You can change the number of entries you see on the screen as well as search for specific information in the dataset.

Show 10 ▾ entries		Search: <input type="text"/>							
mpg ▴ ▾	cyl ▴ ▾	disp ▴ ▾	hp ▴ ▾	drat ▴ ▾	wt ▴ ▾	qsec ▴ ▾	vs ▴ ▾	am ▴ ▾	gear ▴ ▾
21	6	160	110	3.9	2.62	16.46	0	1	4
21	6	160	110	3.9	2.875	17.02	0	1	4
22.8	4	108	93	3.85	2.32	18.61	1	1	4
21.4	6	258	110	3.08	3.215	19.44	1	0	3
18.7	8	360	175	3.15	3.44	17.02	0	0	3
18.1	6	225	105	2.76	3.46	20.22	1	0	3
14.3	8	360	245	3.21	3.57	15.84	0	0	3
24.4	4	146.7	62	3.69	3.19	20	1	0	4
22.8	4	140.8	95	3.92	3.15	22.9	1	0	4
19.2	6	167.6	123	3.92	3.44	18.3	1	0	4

Step 4: Click the **Variables** tab on the left to navigate to a page where you can enter metadata about your variables. See the instructions on the page for details about the information you can add.

DD Creator

☰

1. Project Info

2. Upload Data

3. Variables

4. Category Labels

5. Output

Edit Variable Metadata

—

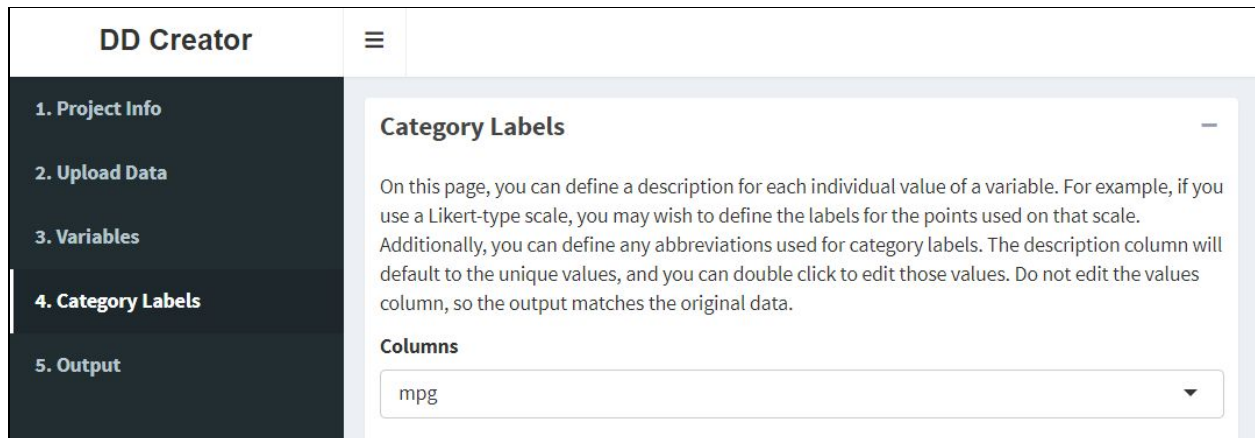
On this page, you will edit the information about the columns/variables in your uploaded dataset. Double click in an individual cell to edit the information provided or add your own information.

- Variable: The name of the column in the dataset. Do not edit.
- Unique Values: This column indicates the number of unique values automatically calculated from your data.
- Missing Values: This column indicates the number of missing values automatically calculated from your data.
- Levels: The levels variable is automatically created for items with less than five unique values. These can be added or deleted for items that should/not include them.
- Description: This column allows you to enter a text description of the variable. Use this column to inform other researchers about the variable, how it is measured, etc. You can define category labels (i.e. for Likert-type scales or abbreviations used in character columns) under step 4 in the left hand menu.

Step 5: Below the instructions, the metadata table will be populated with information from the dataset you uploaded in the previous step. Double-click inside a cell to edit or add information about your variables. The page will save automatically.

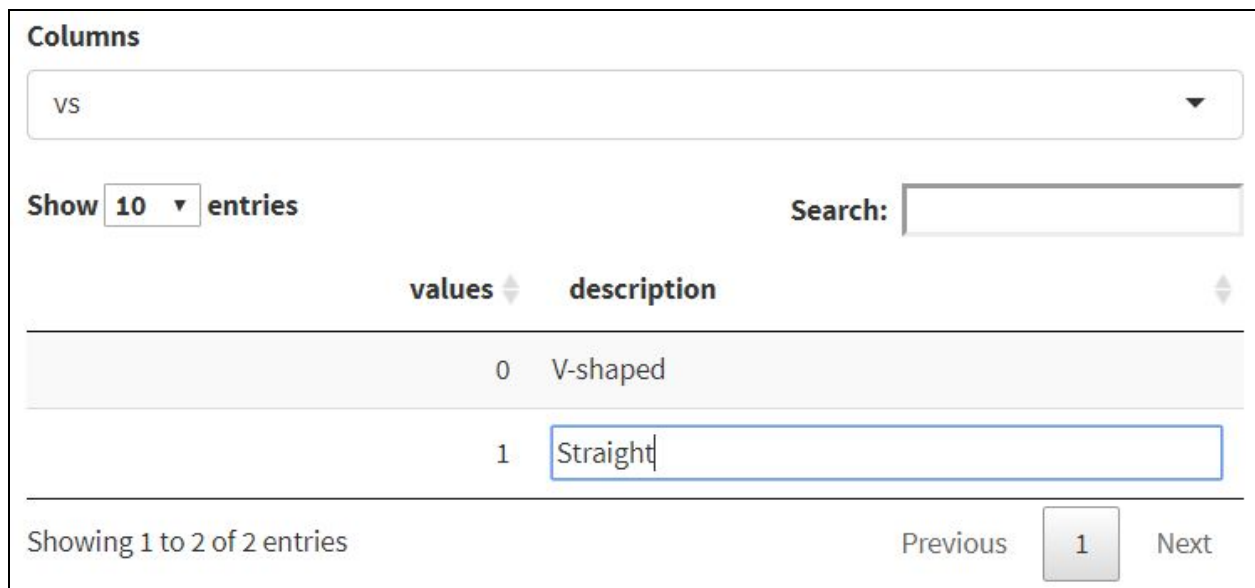
Show 10 entries		Search: <input type="text"/>				
Variable	# Unique Values	# Missing Values	Levels	Description (required)	Type	Minimum Allowable
mpg	25	0		Miles per Gallon	numeric	10.
cyl	3	0	4, 6, 8	Cylinders	integer	
disp	27	0		Displacement (cu. in.)	numeric	71.
hp	22	0		Horsepower	integer	5
drat	22	0			numeric	2.7
wt	29	0			numeric	1.51
qsec	30	0			numeric	14.
vs	2	0	0, 1		integer	

Step 6: Click the **Category Labels** tab on the left to navigate to a page where you can edit category labels for your variables. The **Columns** drop-down list will populate based on your dataset.



The screenshot shows the 'DD Creator' interface. On the left is a sidebar with five tabs: '1. Project Info', '2. Upload Data', '3. Variables', '4. Category Labels' (which is selected and highlighted), and '5. Output'. The main area is titled 'Category Labels' and contains a text block explaining that users can define descriptions for individual values of a variable, such as for a Likert-type scale. Below this text is a 'Columns' section with a drop-down menu currently showing 'mpg'.

Step 7: Use the **Columns** drop-down box to select which variable description to change. The table below will populate based on your selection. Double-click inside a cell to make changes. The page will save automatically.



The screenshot shows the 'Columns' section of the interface. At the top is a drop-down menu with 'VS' selected. Below it are controls for 'Show 10 entries' and a 'Search:' input field. A table with two columns, 'values' and 'description', displays two entries. The first entry has a value of '0' and a description of 'V-shaped'. The second entry has a value of '1' and a description of 'Straight', which is currently being edited in a text box. At the bottom, there is a pagination bar showing 'Showing 1 to 2 of 2 entries', 'Previous' and 'Next' buttons, and a page number '1' in a highlighted box.

values	description
0	V-shaped
1	Straight

Step 8: Click the **Output** tab on the left side of the page. Select the file you would like to download. Variable Metadata and Category Labels will download as CSV files. You may also download an Rdata or JSON file.

DD Creator

☰

1. Project Info

2. Upload Data

3. Variables

4. Category Labels

5. Output

Output

—

On this page, you can download several versions of the information you created in steps 3 and 4.

The Variable Metadata is a CSV file that contains the metadata information you entered on step 3.

The Category Labels file is a CSV file that contains the metadata information for category labels you entered in step 4.

The RData format (.Rdata) is a format designed for use with R, that combines the information you entered in each step into one file with the data and attributes put together.

The JSON format (.Rdata) is a machine readable format that combines the variable and category label metadata into one structured file.

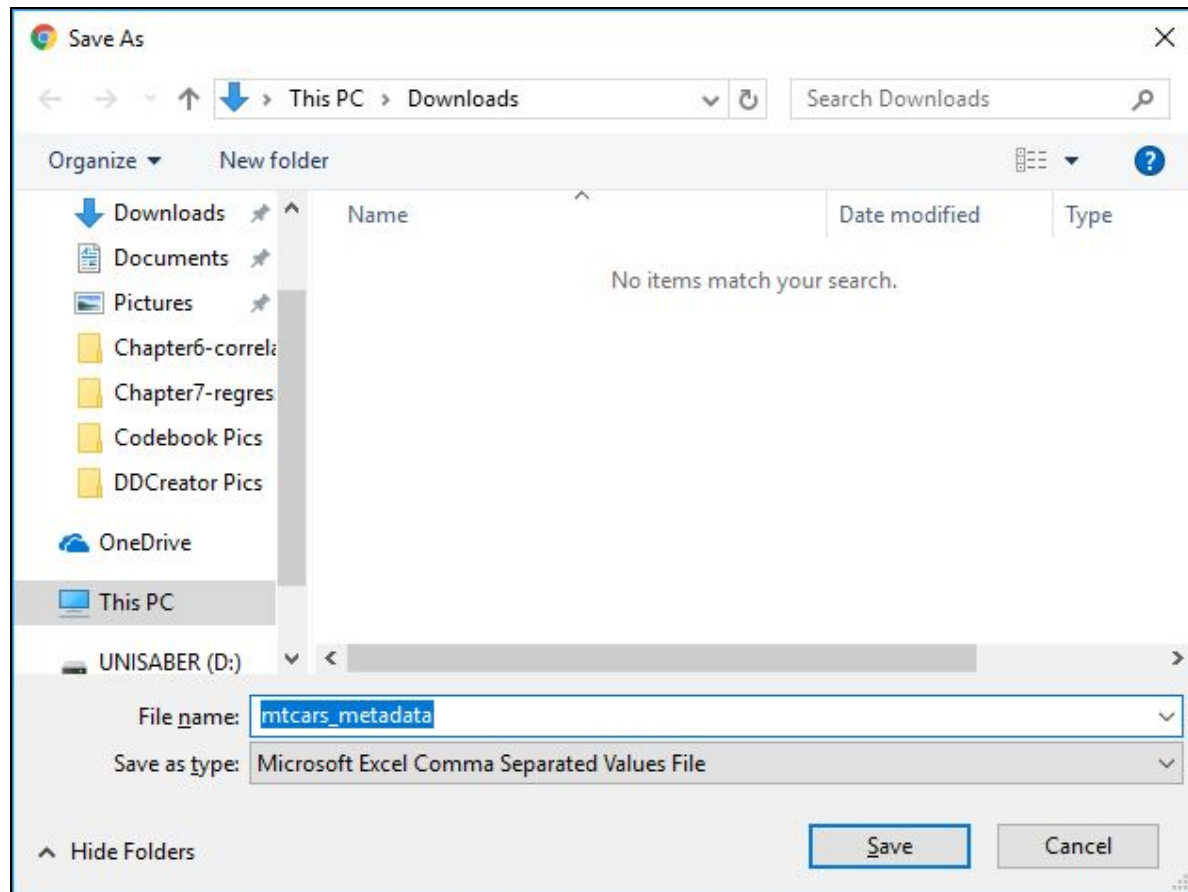
Download Variable Metadata

Download Category Labels

Download Rdata file

Download JSON file

Enter a file name and click **Save**.



Summary

In this tutorial, we provide three accessible options for researchers to create data dictionaries: shareable metadata information that allows others to read and understand the data provided. This type of tutorial is especially critical as transparency practices become more commonplace and as journals and researchers implement guidelines for sharing information and open data. For example, the availability of large, open neuroimaging datasets led to the development of Brain Imaging Data Structure, which defines standards for neurological data (BIDS; Gorgolewski et al., 2016), and a similar movement is occurring in psychology with the psych-DS project (Kline, 2018). This tutorial provides a manageable first step toward generating understandable and reusable metadata for sharing. The applications showcased here will continue to evolve as cohesive standards are formed through group discussion.

Additional Resources

- 1) Boulton, G., Rawlins, M., Vallance, P., & Walport, M. (2011). Science as a public enterprise: the case for open data. *The Lancet*, 377(9778), 1633-1635.
[https://doi.org/10.1016/S0140-6736\(11\)60647-8](https://doi.org/10.1016/S0140-6736(11)60647-8)
- 2) Codebook's website with examples from many data types:
<https://rubenarslan.github.io/codebook/index.html>
- 3) Google Data Explorer: <https://www.google.com/publicdata/directory>

References

- Arslan, R. C. (2018). *How to automatically generate rich codebooks from study metadata*.
<https://doi.org/10.31234/osf.io/5qc6h>
- Boettiger, C., Chamberlain, S., Fournier, A., Hondula, K., Krystalli, A., Mecum, B., ... Woo, K. (2018, October 10). *dataspice*. Retrieved November 24, 2018, from
<https://github.com/ropenscilabs/dataspice>
- Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2018). *shiny: Web Application Framework for R* (Version R package shiny version 1.2.0). Retrieved from Comprehensive R Archive Network (CRAN) website: <https://CRAN.R-project.org/package=shiny>
- DeBruine, L., Buchanan, E. M., & Mohr, A. H. (2018, July 1). *ddcreator*. Retrieved November 24, 2018, from <https://github.com/debruine/ddcreator>
- Gorgolewski, K. J., Auer, T., Calhoun, V. D., Cameron Craddock, R., Das, S., Duff, E. P., ... Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3, 160044.
<https://doi.org/10.1038/sdata.2016.44>
- Hardwicke, T. E., Mathur, M. B., MacDonald, K. E., Nilsson, G., Banks, G. C., Kidwell, M., ... Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal *Cognition*. *Royal Society Open Science*, 5(8). <https://doi.org/10.1098/rsos.180448>
- Henderson, H. V., & Velleman, P. F. (1981). Building Multiple Regression Models Interactively. *Biometrics*, 37(2), 391–411. <https://doi.org/10.2307/2530428>
- Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V. M., Nichols, T. E., &

- Wagenmakers, E.-J. (2018). Data Sharing in Psychology: A Survey on Barriers and Preconditions. *Advances in Methods and Practices in Psychological Science*, 1(1), 70–85.
<https://doi.org/10.1177/2515245917751886>
- Kline, M. (2018). psych-DS. Retrieved from <https://github.com/psych-ds/psych-DS>
- McDaniel, G., & International Business Machines Corporation. (1994). *IBM Dictionary of Computing*. Retrieved from
https://books.google.com/books/about/IBM_Dictionary_of_Computing.html?hl=&id=j9ZjQgAACAAJ
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., ... Yarkoni, T. (2016). How open science helps researchers succeed. *eLife*, 5.
<https://doi.org/10.7554/eLife.16800>
- Noy, N. (2018, September 5). Making it easier to discover datasets. Retrieved November 24, 2018, from Google website:
<https://www.blog.google/products/search/making-it-easier-discover-datasets/>
- Rouder, J. N. (2016). The what, why, and how of born-open data. *Behavior Research Methods*, 48(3), 1062–1069. <https://doi.org/10.3758/s13428-015-0630-z>
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Specification Curve: Descriptive and Inferential Statistics on All Reasonable Specifications. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.2694998>
- Vadillo, M. A., Gold, N., & Osman, M. (2018). Searching for the bottom of the ego well: failure to uncover ego depletion in Many Labs 3. *Royal Society Open Science*, 5(8), 180390.
<https://doi.org/10.1098/rsos.180390>

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10), 1–23.

<https://doi.org/10.18637/jss.v059.i10>

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., ...

Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>