Getting started creating data dictionaries

Document Version:
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

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

Erin M. Buchanan[1][2]

Sarah E. Crain[1]

Ari L. Cunningham[1]

Hannah R. Johnson[1]

Hannah Stash[1]

Marietta Papadatou-Pastou[3]

Peder Mortvedt Isager[4]

Rickard Carlsson[5]

Balazs Aczel[6]

[1]Missouri State University

[2]Harrisburg University of Science and Technology

[3]National and Kapodistrian University of Athens

[4]Eindhoven University of Technology

[5]Linnaeus University

[6]Eötvös Loránd University

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

Disclosures:

Abstract

As researchers embrace open and transparent data sharing, they will need to provide information about their data that effectively helps others understand its contents. Without proper documentation, data stored in online repositories such as OSF will often be rendered unfindable and unreadable by other researchers and indexing search engines. Data dictionaries and codebooks provide a wealth of information about variables, data collection, and other important facets of a dataset. This information, called metadata, provides key insights into how the data might be further used in research and facilitates search engine indexing to reach a broader audience of interested parties. This tutorial first explains the terminology and standards surrounding data dictionaries and codebooks. We then present a guided workflow of the entire process from source data (e.g., survey answers on Qualtrics) to an openly shared dataset accompanied by a data dictionary or codebook that follows an agreed-upon standard. Finally, we explain how to use freely available web applications to assist this process of ensuring that psychology data are findable, accessible, interoperable, and reusable (FAIR; Wilkinson et al., 2016).

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

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. Publications with open datasets have higher citation rates compared to papers that do not have open datasets (McKiernan et al., 2016). Open data further allows scientists to develop and test new hypotheses (e.g., Vadillo, Gold, & Osman, 2018; the Human Connectome Project; Van Essen et al., 2013), 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). Despite these benefits, there are no set standards for making data public (Hardwicke et al., 2018; Houtkoop et al., 2018). One concern is that the shared data are not reusable without some meta-level description of the contents of the dataset (e.g. the meaning of variable names, the meaning of factor levels, details about measurement scales used, etc.). Further, open data may not be findable without also making available, in a machine-readable and searchable format, corresponding metadata that describe the dataset.

A data dictionary is a supplementary document that details the information provided in a dataset, often including the meaning and attributes of the included variables as well as information about the creation, format, and usage of the data (McDaniel & International Business Machines Corporation, 1994). Data dictionaries can be contrasted with codebooks, which are used to denote survey data, while the former include information about the data file structure. Data dictionaries may include the information provided in a codebook and vice versa,

and sometimes these terms are used interchangeably (University of Iowa Libraries, n.d.). For

both document types, the information provided about the data is called metadata.

This tutorial will demonstrate two simple procedures for non-programmers to create a

codebook and data dictionary that describes research data in the social sciences. By following

these procedures, a researcher can begin to create metadata with the goal of sharing the files on a

platform for others to read.

Metadata Format

In order to provide open data, researchers should prepare both human- (i.e., researcher)

and machine-readable metadata in the form of a codebook or data dictionary. Human-readable

data may include a descriptive report of the variables included in the data, summary of the

project, or data collection dates provided in text format. In contrast, machine-readable formats

are designed to allow computers to easily process the data, which requires data to be structured

in a specific and standardized way. A simple example of a machine-readable format is a barcode,

which is structured to provide data to a computer when scanned. Without the structure of a

machine-readable format, it would be difficult for computers, and hence search engines, to

automatically process information contained within those documents.

Two data formats that purport to be both human- and machine-readable are eXtensible

Markup Language (XML) and JavaScript Object Notation (JSON). XML is often used to embed

metadata into Word and PDF documents to save the author information, creation dates, digital

object identifiers, and more. JSON is often used for providing structured metadata for web

purposes because it is considered "lightweight" (i.e., simply structured for quick and easy

processing; Crockford, n.d.). JSON files are formatted in the style of a dictionary. Each entry

includes a definition stored as name-value pairs. The following JSON code provides an example

of how you might provide metadata about authors of a project:

```
{
  "author":[
    {"firstName":"Erin", "lastName":"Buchanan"},
    {"firstName":"Sarah", "lastName":"Crain"},
    {"firstName":"Ari", "lastName":"Cunningham"}
  ]
}
```

The name entry `author` is defined with three values (i.e., three authors of this paper). The

names of the authors are then broken into smaller name-value pairs, `firstName` and `lastName`

along with their respective values (`Erin, Sarah, Ari` and `Buchanan, Crain,`

`Cunningham`). A newer form of JSON, JSON-Linked Data (JSON-LD), should be primarily

used for sharing metadata. The LD format was designed as part of the Resource Descriptive

Framework specifically for metadata (Brickley, Carroll, Klyne, & Parsia, 2004). This version of

JSON includes `context` and `type` information that links JSON name-value pairs into a formal

representation. An example of JSON-LD using data from the Semantic Priming Project

(Hutchison et al., 2013) might include the following:

```
{
  "@context": ["https://schema.org/"],
  "@type": ["Dataset"],
  "name": ["The Semantic Priming Project"],
  "fileFormat": [".csv"],
  "contentUrl": ["http://spp.montana.edu/"]
}
```

The `context` provides the reference for the standards or structure of identifying information that will be used in the file, while the `type` identifies the specific scheme. Schema.org is a collaborative group of individuals that work as a community to create a "shared vocabulary" that allows machine-readable formats to be interpreted consistently across different instances ("About Schema.org," n.d.). For the purposes of metadata creation, the Dataset schema provides a formatting guide for the expected name-value pairs and embedded types that might be present in a dataset. For example, authors are embedded in a `person` type:

```
{
   "author":[
     {
        "@type":["Person"],
        "identifier":["https://orcid.org/0000-0002-9689-4189"],
        "givenName":["Erin"],
        "familyName":["Buchanan"],
        "email":["ebuchanan@harrisburgu.edu"],
        "affiliation":["Harrisburg University of Science and Technology"]
     }
   ]
}
```

By using JSON-LD paired with Schema.org types, we can create a metadata file that provides a wealth of readable, consistent information for others to use. The `variableMeasured` option for datasets can be structured to detail each measured outcome in a dataset:

```
{
   "variableMeasured":[
     {
        "@type":["PropertyValue"],
        "identifier":["Q1_3"],
        "unitText":["integer"],
        "minValue":["1"],
        "maxValue":["6"],
```
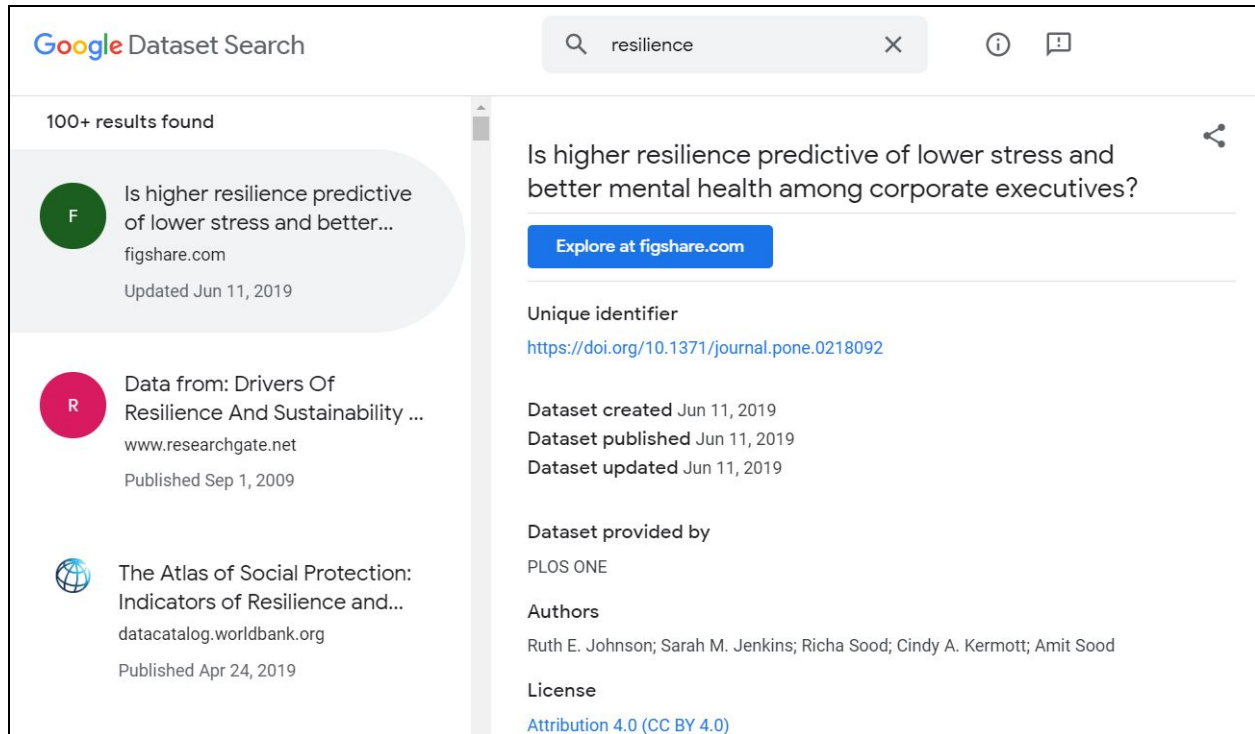
```
      "description":["IN THE LAST TWO WEEKS - I was attentive and
aware of my emotions"]
    }
  ]
}
```

The framework provided by Schema.org can be extended by individual research communities. For example, https://bioschemas.org/ focuses on extending new types and properties for data relevant to the life sciences, while the Brain Imaging Data Structure (BIDS; Gorgolewski et al., 2016) provides structure specifically for brain imaging data. Psych-DS represents the current effort to expand schemas for psychological data (Kline, 2018), and readers are encouraged to join the online community.
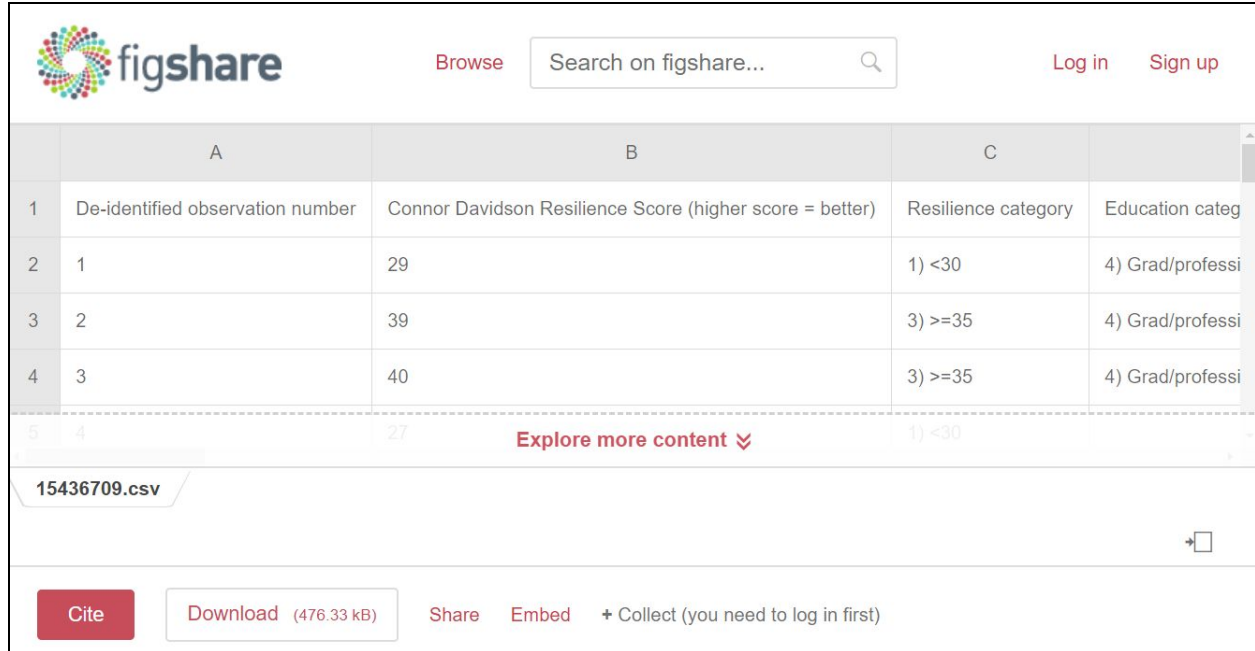
An additional advantage of the JSON-LD and Schema.org framework is the ability to index data in a searchable portal, as these formats are optimized for search engines. Google has launched Dataset Search to enable researchers and others to find data that has been published online (https://toolbox.google.com/datasetsearch; Noy, 2018). Their guidelines for dataset discovery include using JSON-LD and Schema.org compliant formatting. The benefit of indexing to researchers who wish to find datasets cannot be understated. Figure 1 portrays the use of the Google Dataset Search engine to find a dataset related to "resilience". The first record identifies a published paper (Kermott, Johnson, Sood, Jenkins, & Sood, 2019) that has been shared on figshare.com. Clicking "Explore at figshare.com" links to the dataset with embedded metadata on the figshare website, as depicted in Figure 2. The metadata for this dataset help clarify citation information, variables measured, value labels for continuous measures (i.e., 10 - As good as it can be for overall quality of life), and direction of scores (i.e., higher is better).

With this information, the data can be used by interested researchers to reproduce the results

from the study, extend data for testing new hypotheses, assess sample size and power for

planning new studies, and conduct meta-analyses, among other uses.

Figure 1. Results from searching for "resilience" on Google Dataset Search.

Figure 2. Kermott et al. (2019) dataset shared on figshare.com.

Metadata Requirements

What makes a dataset minimally readable? For datasets in psychology, minimal

information likely includes basic bibliographic information (author names, publication date,

DOI, etc.) and a detailed description of the information provided in each variable of the dataset

(i.e., each column of the tabular datafile). Variable-specific information could include the type of

data in the variable (numbers, character strings, etc.), missing value denominators (NA, 999, "",

etc.), minimum values, maximum values, a description of the variable (i.e., questionnaire,

variable name, item number, etc.), and a mapping of value labels to numeric data when

appropriate (i.e., 1 = Strongly Agree, 5 = Strongly Disagree) (Bowman, n.d.; Smithsonian

Libraries, 2018). The wide variety of types of data available in the behavioral sciences limits the

ability to create a catch-all set of coding criteria. A general rule of thumb states that metadata

should be understandable for any question one might have about the data (Moellering, Aalders,

& Crane, 2005).

Metadata Creation

Figure 3 presents a flowchart of the metadata creation process. The left side starts with

the rules or structure one should follow for developing a machine-readable data dictionary. Next,

the data are converted to a data dictionary or codebook by using a tool, like Codebook or Data

Dictionary Creator, that creates the metadata output in JSON-LD and/or HTML format. Finally,

these are stored in an online repository—such as the Open Science Framework (OSF), GitHub,

Figshare, or Zenodo—to share with a larger audience. On our OSF page (https://osf.io/3y2ex/),

we have included a multipart tutorial that helps readers walk through the "tools for the rules" on

creating metadata for online sharing. In part one, we demonstrate how to export data from an

online platform, Qualtrics (Qualtrics, 2019), and explain how to maintain the metadata provided

automatically by the survey software. In the second section, we include videos that demonstrate

how to create a codebook and data dictionary from the downloaded data. The suggested

applications are described below. The final section describes how to upload and share your data

and metadata to an online platform. We also demonstrate Google Dataset Search to help

researchers who want to find existing data in their respective areas.

Input

Dataset

Structure

Google
Schema.org

Tools

Codebook
DD Creator

Share

OSF
GitHub
Figshare
Zenodo

Output

JSON-LD
HTML
Metadata
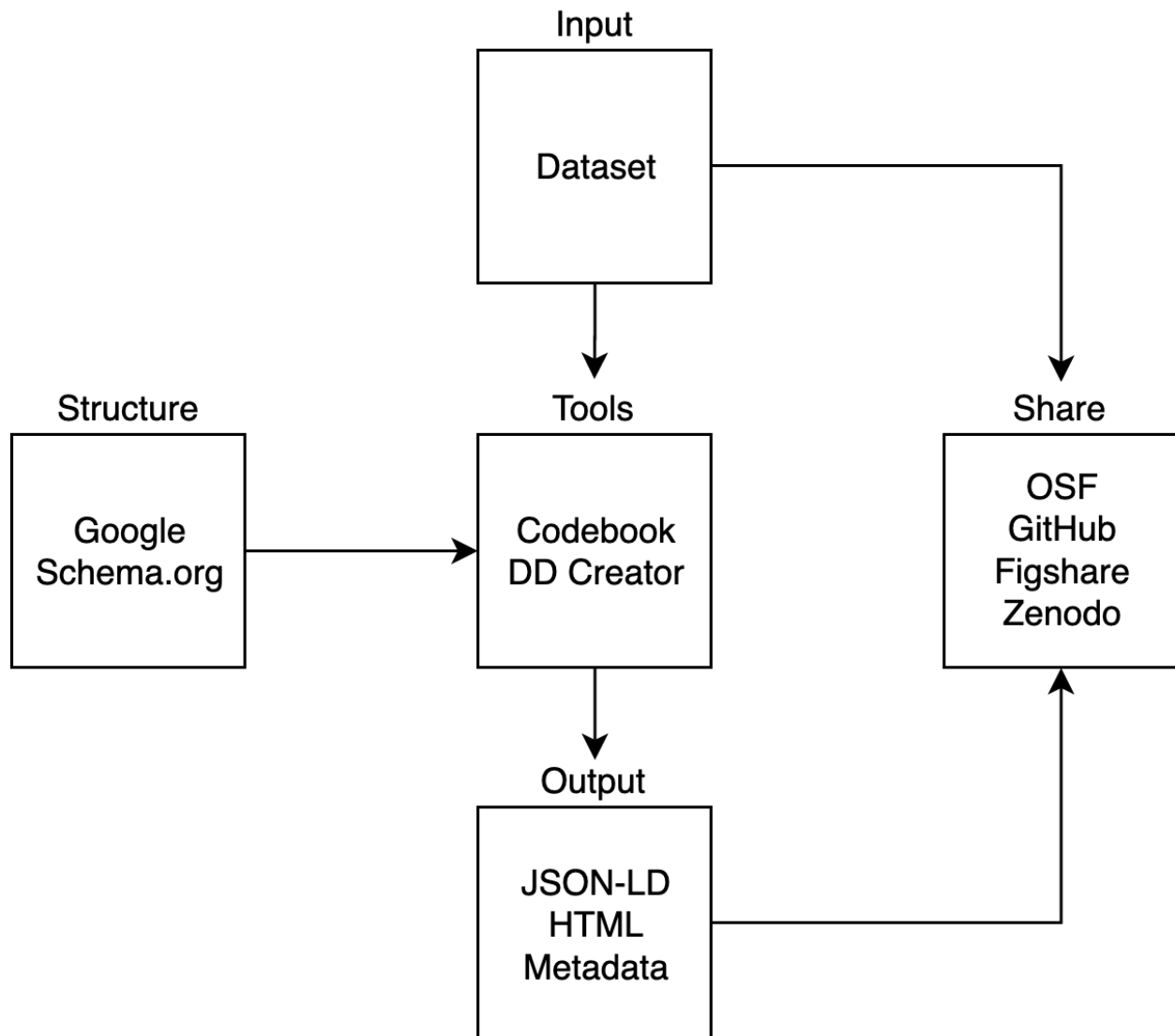
Figure 3. A flowchart indicating the process of creating a data dictionary or codebook. The left side indicates the rules for creating a data dictionary or codebook, which are programmed into the suggested applications.. The middle column denotes how data is processed through the application to create the appropriate JSON-LD and/or HTML file. The last column shows the final step of creating open data, which is sharing the dataset and data dictionary/codebook in an online platform.

The Applications

Table 1 summarizes the properties and relative benefits of the applications *Codebook*[1]

and Data Dictionary Creator. We have also included detailed video tutorials as supplementary

material to be updated as the applications update, and the links are provided online at

https://osf.io/3y2ex/. A wide range of data formats can be uploaded into these applications,

including SPSS/SAS, comma separated values (CSV), plain text, and Excel file formats. In both

applications, data is imported using the *rio* package in *R*, which allows for a large number of

formatted data types[2] (Chan, Chan, Leeper, & Becker, 2018). The output from these applications

includes HTML with embedded JSON-LD, comma separated values (CSV), standalone

JSON-LD, and Rdata with embedded attributes. Once the data dictionary is created, these files

can be shared alongside the dataset in the same folder of a web repository (see Rouder, 2016 for

a tutorial). In the case of multiple datasets and dictionaries, separate subfolders or naming cues

should be included to ensure that researchers can map each dataset to the appropriate dictionary

or codebook.

---

[1] *Codebook* in italics refers to the application and *R* program.
[2] You can view the data options here:
https://cran.r-project.org/web/packages/rio/vignettes/rio.html#supported_file_formats

Table 1.

Information about the Benefits for each Application

	Codebook	Data Dictionary Creator
Citation	Arslan (2019)	Buchanan, De Bruine, & Mohr (2019)
Interface	Web Application *R* package	Web Application
Link	https://codebook.formr.org	https://doomlab.shinyapps.io/ddcreator/
Input	Nearly all formatted data	Nearly all formatted data
Output	HTML report containing embedded JSON-LD, CSV, separate JSON-LD	CSV files of metadata, JSON-LD, Rdata
Benefits	Easiest to use Quick metadata generation Generates a summary for each variable in a readable format Uses embedded metadata	Specifies a separate section for category labels Rdata output Detailed data entry options for non *R* users Uses embedded metadata

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 example dataset contains a few demographic questions (gender, race), the 14-Item Resilience Scale (RS-14; Wagnild, 2009), the Meaning in Life Questionnaire (Steger, Frazier, Oishi, & Kaler, 2011), and part of the Multidimensional Psychological Flexibility Inventory (MPFI; Rolffs, Rogge, & Wilson, 2018). This data was presented as part of a workshop on data quality indicators (Buchanan & Azevedo, 2019) to demonstrate how to assess Likert-style data in surveys. Other variables include page timing, click counts, and a few attention check measures. In general, the requirements for data in these tutorials is that they: 1) be in a file format that is readable by one of the demonstrated

applications as shown in Table 1, 2) include participants[3], and 3) include participant data in the

form of variables. Data should/may be arranged according to tidy data principles: 1) each

variable is included in its own column, 2) each observation is its own row, and 3) each value is

represented in its own cell (Wickham, 2014).

Codebook

       Codebook (Arslan, 2019) is an *R* package with a corresponding website that allows

researchers to create reports of their data, including reliability statistics (i.e., alpha) and

summaries of items (histograms, descriptive statistics). Metadata embedded in the datafile (such

as item labels) are automatically included in the report. This tutorial focuses on the web interface

version of *Codebook*, as shown in Figure 4. *R* users are encouraged to consult Arslan (2019) for

a complete *R* tutorial. *Codebook'*s web interface is simple and easy to use, and automatically

imports embedded metadata that are provided in popular statistical software (e.g., SPSS, SAS).

The output from *Codebook* includes an HTML report with embedded JSON-LD to ensure data

can be indexed in Google Dataset Search; thus, the output is both human- and machine-readable.

The data and codebook created from this web application could then be shared on sites like those

shown in Figure 3. The online web application is best for researchers who have data files with

embedded metadata, as the ability to edit and add information is limited.

---

[3] A participant is defined as a person/animal from a study or a unit from another type of studied data, such as stimuli, words, etc.
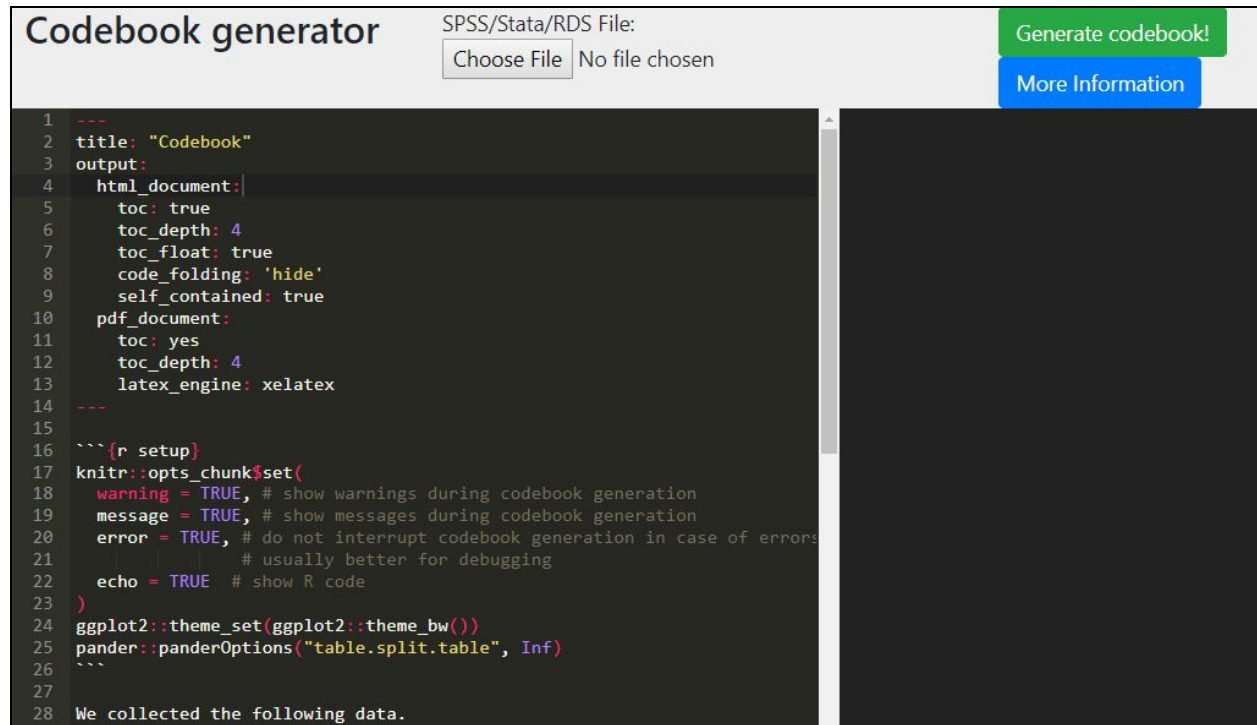
Figure 4. The web application interface for *Codebook*.

Data Dictionary Creator

Data Dictionary Creator breaks down the metadata entry into five steps, as shown in the

left side of Figure 5. First, the user will upload the data file for processing only (i.e., the data is

not stored permanently), and they can preview the uploaded data to determine if it was imported

correctly. The second step of the process involves entering the metadata for each column

provided in the dataset. The application automatically provides 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. Any embedded metadata from files like

SPSS, SAS, or Qualtrics CSVs (e.g., some metadata is stored in the second row) are imported

into the description or category label attributes. On the third page, category labels can be

provided for both character and numeric data (i.e., Likert-type scales that include labeled

numbers), and these labels can quickly be copied over from one column to an entire scale. On the

fourth page, the user will enter overall project information such as citation, website, funders,

dates of data collection, and authors. Users can then download CSV files of the metadata, a

JSON-LD formatted metadata file, and an Rdata file that includes the dataset and descriptive

information integrated together. This application is built with the *Shiny R* package (Chang,

Cheng, Allaire, Xie, & McPherson, 2019), and the default timeout options (i.e., the amount of

time a user has to interact with an application) were increased to accommodate entering
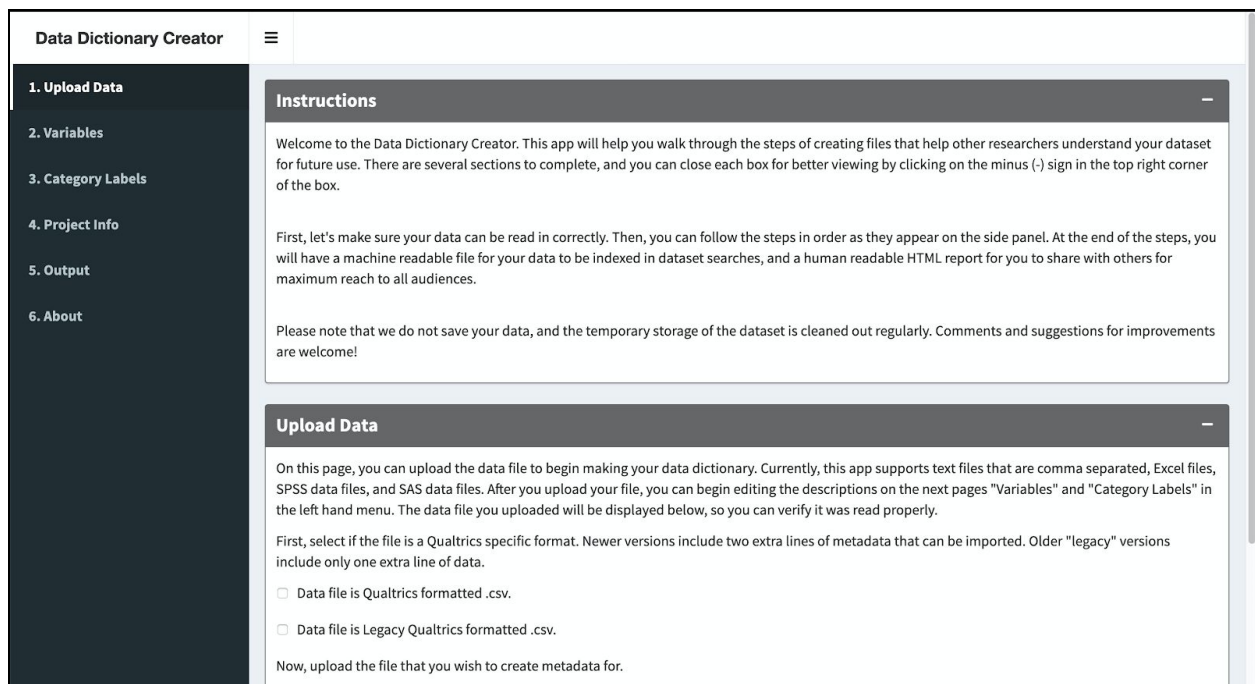
information for complex datasets.

Figure 5. The project information interface for Data Dictionary Creator.

Summary

In this tutorial, we detailed the concepts necessary to understand data dictionaries, codebooks, and metadata and provided a tutorial for researchers to create their own. This type of tutorial is especially critical as transparency practices become more commonplace, and FAIR guidelines are implemented in journals for sharing information and open data. For example, the availability of large, open neuroimaging datasets led to the development of the BIDS, which defines standards for how to curate open neuroscientific data (BIDS; Gorgolewski et al., 2016), and a similar movement is occurring in psychology with the psych-DS project (Kline, 2018). Data may also be published in journals that include data papers, such as Nature Scientific Data, Data in Brief, and Journal of Open Psychology Data. This tutorial provides a manageable first step toward generating understandable and reusable metadata for sharing and publication. The applications showcased here (Gorgolewski et al., 2016) 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

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

About Schema.org. (n.d.). Retrieved December 8, 2019, from schema.org website:

http://schema.org/docs/about.html

Arslan, R. C. (2019). How to automatically document data with the codebook package to

facilitate data reuse. *Advances in Methods and Practices in Psychological Science*, *2*(2),

169–187. https://doi.org/10.1177/2515245919838783

Bowman, S. (n.d.). How to make a data dictionary. Retrieved December 8, 2019, from OSF

Guides website:

http://help.osf.io/hc/en-us/articles/360019739054-How-to-Make-a-Data-Dictionary

Brickley, D., Carroll, J., Klyne, G., & Parsia, B. (2004, February 10). RDF/XML syntax

specification (revised). Retrieved December 8, 2019, from W3C Recommendations website:

https://www.w3.org/TR/REC-rdf-syntax/

Buchanan, E. M., & Azevedo, F. (2019, July). *Workshop - Statistics are useless without suitable*

data: How to implement and assess for data quality. Presented at the Society for the

Improvement of Psychological Science (SIPS) 2019 Meeting , Rotterdam, The Netherlands.

https://doi.org/10.17605/OSF.IO/X53KN

Buchanan, E. M., DeBruine, L., & Mohr, A. H. (2019). Data Dictionary Creator (Version 0.1.1)

[R]. Retrieved from https://github.com/doomlab/dd-creator/

Chan, C.-H., Chan, G. C. H., Leeper, T. J., & Becker, J. (2018). rio: A Swiss-army knife for data

file I/O (Version R package version 0.5.16). Retrieved from

https://cran.r-project.org/package=rio

Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2019). shiny: Web application

framework for R (Version 1.4.0). Retrieved from https://cran.r-project.org/package=shiny

Crockford, D. (n.d.). Introducing JSON. Retrieved December 8, 2019, from JSON website: https://www.json.org/json-en.html

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., Nilsonne, 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), 180448. https://doi.org/10.1098/rsos.180448

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

Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C.-S., … Buchanan, E. M. (2013). The semantic priming project. *Behavior Research Methods*, *45*(4), 1099–1114. https://doi.org/10.3758/s13428-012-0304-z

Kermott, C. A., Johnson, R. E., Sood, R., Jenkins, S. M., & Sood, A. (2019). Is higher resilience predictive of lower stress and better mental health among corporate executives? *PloS One*, *14*(6), e0218092. https://doi.org/10.1371/journal.pone.0218092

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=j9ZjQ gAACAAJ

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

Moellering, H., Aalders, H. J., & Crane, A. (2005). *World spatial metadata standards: Scientific and technical characteristics, and full descriptions with crosstable*. Retrieved from https://play.google.com/store/books/details?id=cDdKL634FKwC

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/

Qualtrics. (2019). Qualtrics. Retrieved December 5, 2019, from Qualtrics website:

https://www.qualtrics.com

Rolffs, J. L., Rogge, R. D., & Wilson, K. G. (2018). Disentangling components of flexibility via the Hexaflex Model: Development and validation of the Multidimensional Psychological Flexibility Inventory (MPFI). *Assessment*, *25*(4), 458–482.

https://doi.org/10.1177/1073191116645905

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

Smithsonian Libraries. (2018). *Smithsonian data management best practices*. Retrieved from

https://library.si.edu/sites/default/files/tutorial/pdf/datadictionaries20180226.pdf

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2011). Meaning in Life Questionnaire.

PsycTESTS Dataset. https://doi.org/10.1037/t01074-000

University of Iowa Libraries. (n.d.). Readme, data dictionaries, codebooks. Retrieved December

8, 2019, from Research Data Services website:

https://www.lib.uiowa.edu/data/manage/documenting/readme/

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

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K., &

WU-Minn HCP Consortium. (2013). The WU-Minn Human Connectome Project: An

overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

Wagnild, G. M. (2009). *The Resilience Scale user's guide: For the U.S. English version of the

Resilience Scale and the 14-Item Resilience Scale (RS-14)*. Retrieved from

https://books.google.com/books/about/The_Resilience_Scale_User_s_Guide.html?hl=&id=

A_7VSAAACAAJ

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