



1-1-2021

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

Erin M. Buchanan
Missouri State University

Sarah E. Crain
MSU Graduate Student

Arielle L. Cunningham
MSU Graduate Student

Hannah R. Johnson
MSU Graduate Student

Hannah Stash
MSU Graduate Student

See next page for additional authors

Follow this and additional works at: <https://bearworks.missouristate.edu/articles-chhs>

Recommended Citation

Buchanan, Erin M., Sarah E. Crain, Ari L. Cunningham, Hannah R. Johnson, Hannah Stash, Marietta Papadatou-Pastou, Peder M. Isager, Rickard Carlsson, and Balazs Aczel. "Getting started creating data dictionaries: How to create a shareable data set." *Advances in Methods and Practices in Psychological Science* 4, no. 1 (2021): 2515245920928007.

This article or document was made available through BearWorks, the institutional repository of Missouri State University. The work contained in it may be protected by copyright and require permission of the copyright holder for reuse or redistribution.

For more information, please contact BearWorks@library.missouristate.edu.

Authors

Erin M. Buchanan, Sarah E. Crain, Arielle L. Cunningham, Hannah R. Johnson, Hannah Stash, Marietta Papadatou-Pastou, Peder M. Isager, Rickard Carlsson, and Balazs Aczel

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



**Erin M. Buchanan^{1,2} , Sarah E. Crain¹, Ari L. Cunningham¹,
Hannah R. Johnson¹, Hannah Stash¹, Marietta Papadatou-Pastou³,
Peder M. Isager⁴ , Rickard Carlsson⁵, and Balazs Aczel⁶**

¹Department of Psychology, Missouri State University; ²Analytics, Harrisburg University of Science and Technology; ³Department of Special Education and Psychology, National and Kapodistrian University of Athens; ⁴Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology; ⁵Department of Psychology, Linnaeus University; and ⁶Institute of Psychology, ELTE Eötvös Loránd University

Advances in Methods and
Practices in Psychological Science
January-March 2021, Vol. 4, No. 1,
pp. 1–10
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2515245920928007
www.psychologicalscience.org/AMPPS



Abstract

As researchers embrace open and transparent data sharing, they will need to provide information about their data that effectively helps others understand their data sets' 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 data set. 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 terminology and standards relevant to data dictionaries and codebooks. Accompanying information on OSF presents a guided workflow of the entire process from source data (e.g., survey answers on Qualtrics) to an openly shared data set accompanied by a data dictionary or codebook that follows an agreed-upon standard. Finally, we discuss freely available Web applications to assist this process of ensuring that psychology data are findable, accessible, interoperable, and reusable.

Keywords

reproducibility, metadata, data dictionary, codebook, open materials

Received 6/6/19; Revision accepted 4/7/20

Open data sets are beneficial for both individual researchers and the scientific community as a whole. Articles with open data sets reach more researchers, and thus convey their findings to a wider audience. Publications with open data sets have higher citation rates compared with papers that do not have open data sets (McKiernan et al., 2016). Open data further allow scientists to develop and test new hypotheses (e.g., Vadillo et al., 2018; the Human Connectome Project—Van Essen et al., 2013), investigate multiple analytic perspectives by applying them to different data sets (e.g., Simonsohn et al., 2015), and, importantly, identify and correct errors that would otherwise create noise in the literature (Piwovar & Vision, 2013). The 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 shared data are not reusable without some meta-level description of the contents of the data set (e.g., the meaning of variable names, the meaning of factor levels, details about measurement scales used). Further, open data may not be findable if the corresponding metadata that describe the data set are not available in a machine-readable and -searchable format.

Corresponding Author:

Erin M. Buchanan, Analytics, Harrisburg University of Science and Technology
E-mail: ebuchanan@harrisburgu.edu



Creative Commons NonCommercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits noncommercial use, reproduction, and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

A data dictionary is a supplementary document that details the information provided in a data set. Data dictionaries usually include the meaning and attributes of the contained 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 customarily used to describe survey data, but do not additionally include information about the data-file structure, as data dictionaries do. These terms are often used interchangeably, as data dictionaries may include a codebook; however, data dictionaries provide a complete picture of the shared data set (University of Iowa Libraries, n.d.). For both document types, the information provided about the data is called metadata. This Tutorial and the accompanying online materials demonstrate two applications that nonprogrammers can use to create codebooks and data dictionaries that describe research data in the social sciences, with the goal of sharing files on a platform for other researchers to read.

Disclosures

The materials for this Tutorial can be found at <https://osf.io/3y2ex/>. These materials include detailed video tutorials that will be updated as the demonstrated applications are updated. The code for the *Data Dictionary Creator* application can be found at <https://github.com/doomlab/dd-creator/>.

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 data dictionary, with included codebook if necessary. Human-readable data may include a descriptive report of the variables included in the data, a 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 the data to be structured in a specific and standardized way. A simple example of a machine-readable format is the format for a bar code, 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 a data dictionary.

Two data formats that are purported 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 author information, creation dates, digital

object identifiers (DOIs), 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 is an example of how you might provide metadata about the 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 article). The names of the authors are separated 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 specifically for metadata as part of the Resource Descriptive Framework (RDF Core Working Group, 2004). This version of JSON includes context and type information that links JSON name-value pairs into a formal representation. Following is an example of JSON-LD using data from the Semantic Priming Project (Hutchison et al., 2013):

```
{
  "@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 the identifying information that will be used in the file, and the `type` identifies the specific scheme. Schema.org is a collaborative group of individuals who 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 data set. 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, you can create a metadata file that provides a wealth of readable, consistent information for other researchers to use. The `variableMeasured` option for data sets can be structured to detail each measured outcome in a data set. The following example is the code for a survey question, `Q1_3`, that ranges in values from 1 to 6:

```
{
  "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, Bioschemas (<https://bioschemas.org/>) focuses on extending new types and properties for data relevant to the life sciences, and the Brain Imaging Data Structure (BIDS; Gorgolewski et al., 2016) provides structure specifically for brain-imaging data. The psych-DS project represents the current effort to expand schemas for psychological data (Kline, 2018), and we encourage readers to join this 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 other users to find data that have been published online (<https://toolbox.google.com/datasetsearch>; Noy, 2018). Its guidelines for data-set discovery include using

JSON-LD- and Schema.org-compliant formatting. The benefit of indexing to researchers who wish to find data sets cannot be overstated. Figure 1 portrays the use of Dataset Search to find a data set related to “resilience stress.” The first record identifies a published data set (Kermott et al., 2019) that has been shared on figshare.com. Clicking “Explore at figshare.com” links to the data set with embedded metadata on the figshare website, as depicted in Figure 2. The metadata for this data set help clarify citation information, variables measured, value labels for continuous measures (e.g., 10 = as good as it can be for overall quality of life), and direction of scores (e.g., higher is better). With this information, interested researchers can use the data to reproduce the results from the study, test new hypotheses, assess sample size and power for planning new studies, and conduct meta-analyses, among other uses.

Metadata Requirements

What makes a data set minimally readable? For data sets in psychology, minimal information likely includes basic bibliographic information (author names, publication date, DOI, etc.) and a detailed description of the information provided for each variable of the data set (i.e., each column of the tabular data file). Variable-specific information could include the type of data (e.g., numbers, character strings), the missing-value denominator (e.g., NA, 999, “”), the minimum value, the maximum value, a description of the variable (e.g., questionnaire, variable name, item number), and a mapping of value labels to numeric data when appropriate (e.g., 1 = *strongly agree*, 5 = *strongly disagree*; OSF Support, n.d.; Smithsonian Libraries, 2018). The wide variety of types of data available in the behavioral sciences limits the ability to create a catchall set of coding criteria. A general rule of thumb is that metadata should enable users to answer any question they might have about the data (Moellering et al., 2005).

Metadata Creation

Figure 3 presents a flowchart of the metadata-creation and data-sharing 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 and/or codebook by using a tool, such as *codebook*¹ (Arslan, 2019) or *Data Dictionary Creator* (Buchanan et al., 2019), that creates the metadata output in JSON-LD or HTML format. Finally, the data and metadata are stored in an online repository—such as OSF, GitHub, figshare, or Zenodo—to share with a larger audience. On our OSF page (<https://osf.io/3y2ex/>) that accompanies this article, we have included a multi-part tutorial that will help you walk through the “tools for

resilience stress

X

Last updatedDownload formatUsage rightsTopicFree

100+ datasets found

Data from: Self-Guided Multimedia Stress Management and Resilience Training

catalog.data.gov
data.nasa.gov

Updated Feb 28, 2019

Is higher resilience predictive of lower stress and better mental health...

figshare.com

txt

Updated Jun 11, 2019

Is higher resilience predictive of lower stress and better mental health among corporate executives?

Explore at figshare.com

3 scholarly articles cite this dataset ([View in Google Scholar](#))

txt


Unique identifier
<https://doi.org/10.1371/journal.pone.0218092>

Dataset updated Jun 11, 2019

Dataset provided by
PLOS ONE

Authors

Fig. 1. Results from searching for “resilience stress” on Google Dataset Search.



[Browse](#)

[Log in](#)
[Sign up](#)

	A	B	C	D	E
1	De-identified observation number	Connor Davidson Resilience Score (higher score = better)	Resilience category	Education categorized, excluding	Gender
2	1	29	1) <30	4) Grad/professional school	(1) Male
3	2	39	3) >=35	4) Grad/professional school	(1) Male
4	3	40	3) >=35	4) Grad/professional school	(1) Male

[15436709.csv](#)

pone.0218092.s001.csv (476.33 kB)

[Cite](#)
[Download \(476.33 kB\)](#)
[Share](#)
[Embed](#)
[+ Collect \(you need to log in first\)](#)

MD5: c34c95c176350b27b202c748833395ae

Fig. 2. Screenshot showing a portion of Kernott et al.'s (2019) data set shared on figshare.com.

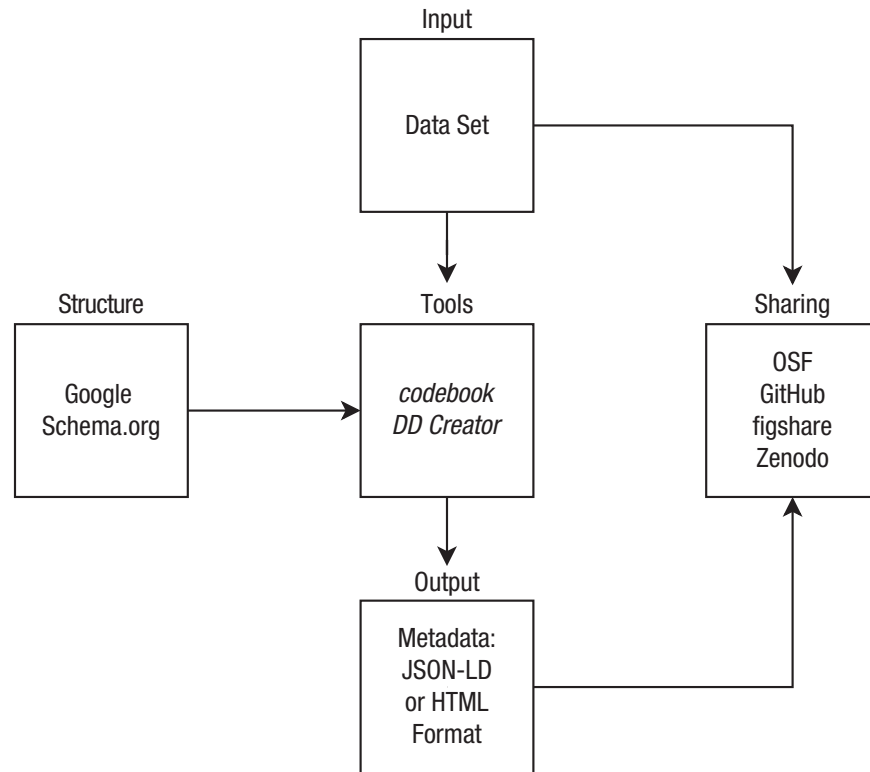


Fig. 3. A flowchart illustrating the process of creating a data dictionary and/or codebook in order to share a data set. The rules for creating a data dictionary or codebook (left side) are programmed into the suggested applications, *codebook* (Arslan, 2019) and *Data Dictionary Creator* (DD Creator; Buchanan et al., 2019). The middle column denotes how data are processed through the selected application to create the metadata, in an appropriate format (JSON-LD or HTML). The right column shows the final step of making the data openly available, which involves sharing the data set and data dictionary or codebook on an online platform.

the rules” on creating metadata for online sharing. In Part 1, we demonstrate how to export data from an online platform, Qualtrics (<https://www.qualtrics.com>), and explain how to maintain the metadata provided automatically by the survey software. In Part 2, videos demonstrate how to create a codebook and a data dictionary from the downloaded data. The suggested applications are described in the following sections of this article. Part 3 of our OSF tutorials describes how to upload and share your data and metadata on an online platform, and we also demonstrate Google Dataset Search to help researchers who want to find existing data in their respective areas.

The Applications

Table 1 summarizes the properties and relative benefits of the *codebook* and *Data Dictionary Creator* applications. We have also provided detailed video tutorials, which can be accessed online at <https://osf.io/3y2ex/>. Data in a wide range of formats, including SPSS or SAS, comma separated values (csv), plain text, and Excel

formats, can be uploaded into these applications. In both applications, data are imported using the *rio* package (Chan et al., 2018) in R, which supports numerous data types.² The output from these applications includes HTML with embedded JSON-LD, csv, standalone JSON-LD, and RData with embedded attributes. Once the data dictionary and/or codebook is created, these files can be shared alongside the data set in the same folder of a Web repository (see Rouder, 2016, for a tutorial). In the case of multiple data sets and dictionaries or codebooks, separate subfolders or naming cues should be used to ensure that researchers can map each data set to the appropriate dictionary or codebook.

In the following, we discuss each of the two suggested applications in more detail. The supplementary video tutorials on OSF demonstrate how to use these applications to process a data set and create different types of metadata output. They describe each data-input space and provide examples of possible descriptions of data sets. The example data set contains a few demographic questions (gender, race), the 14-Item Resilience Scale (Wagnild, 2009), the Meaning in Life Questionnaire

Table 1. Comparison of the Two Applications

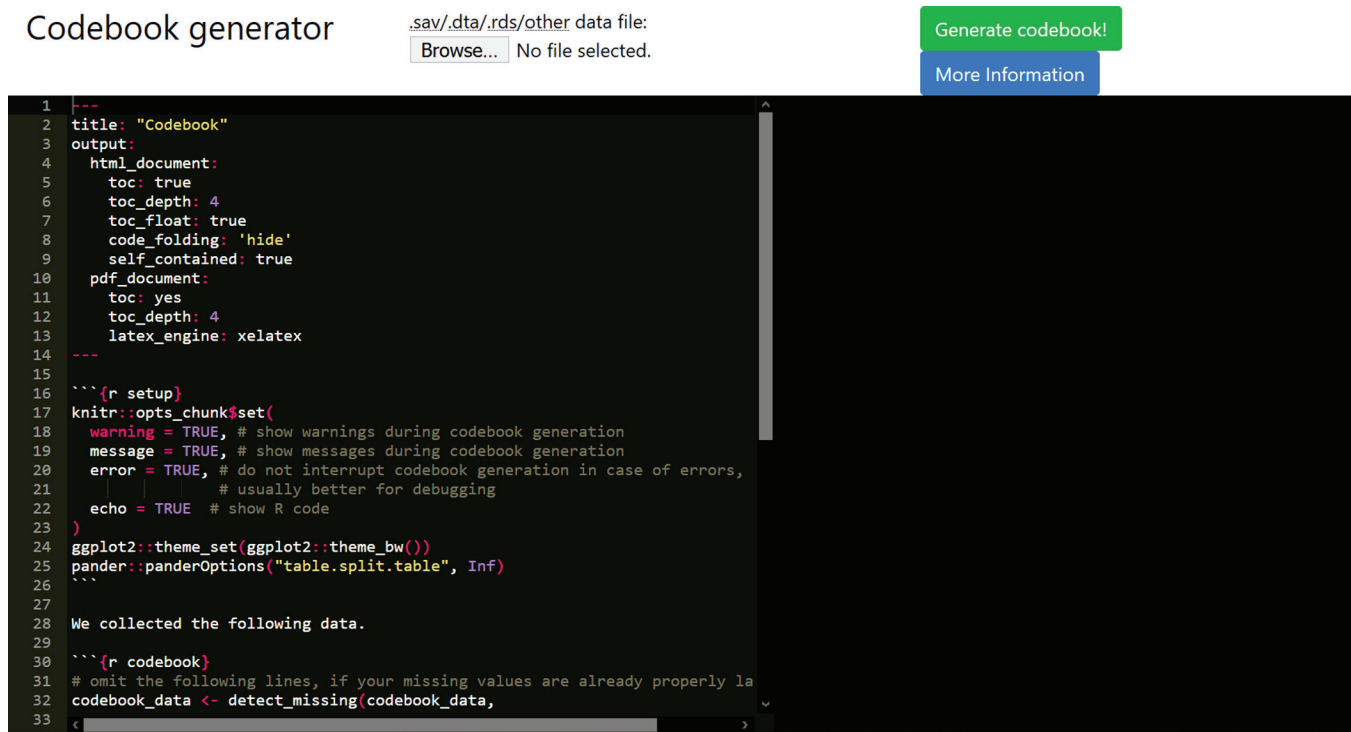
Attribute	<i>codebook</i> (Arslan, 2019)	<i>Data Dictionary Creator</i> (Buchanan et al., 2019)
Interface	Web application, R package	Web application
Link	https://codebook.formr.org	https://doomlab.shinyapps.io/ddcreator/
Input	Nearly all formatted data	Nearly all formatted data
Output	HTML report containing embedded JSON-LD, csv, and separate JSON-LD files	csv files of metadata, JSON-LD-formatted metadata files, RData files
Benefits	Easier to use	Specifies a separate section for category labels
	Generates metadata quickly	Provides RData output
	Generates a summary for each variable in a readable format	Provides detailed data-entry options for non-R users
	Uses embedded metadata	Uses embedded metadata

(Steger et al., 2006), and part of the Multidimensional Psychological Flexibility Inventory (Rolffs et al., 2018). These data were presented as part of a workshop on data-quality indicators (Buchanan & Azevedo, 2019) to demonstrate how to assess Likert-style data using page timing, click counts, and a few attention-check measures. In general, the requirements for the input data are that they (a) be in a file format that is readable by one of the demonstrated applications (see Table 1) and (b) include participant³ data in the form of variables. Data should be arranged according to tidy-data principles, according to which (a) each variable is represented in its own column, (b) each observation is represented in its own

row, and (c) each value is represented in its own cell (Wickham, 2014).

codebook

The *codebook* (Arslan, 2019) R package has a corresponding website (<https://codebook.formr.org>) that allows researchers to create reports of their data, including reliability statistics (e.g., α) and summaries of items (histograms, descriptive statistics). Metadata embedded in the data file (such as item labels) are automatically included in the report. Our videos on OSF focus on the Web interface version of *codebook*, illustrated in Figure 4. We

**Fig. 4.** Screenshot of *codebook*'s Web interface.

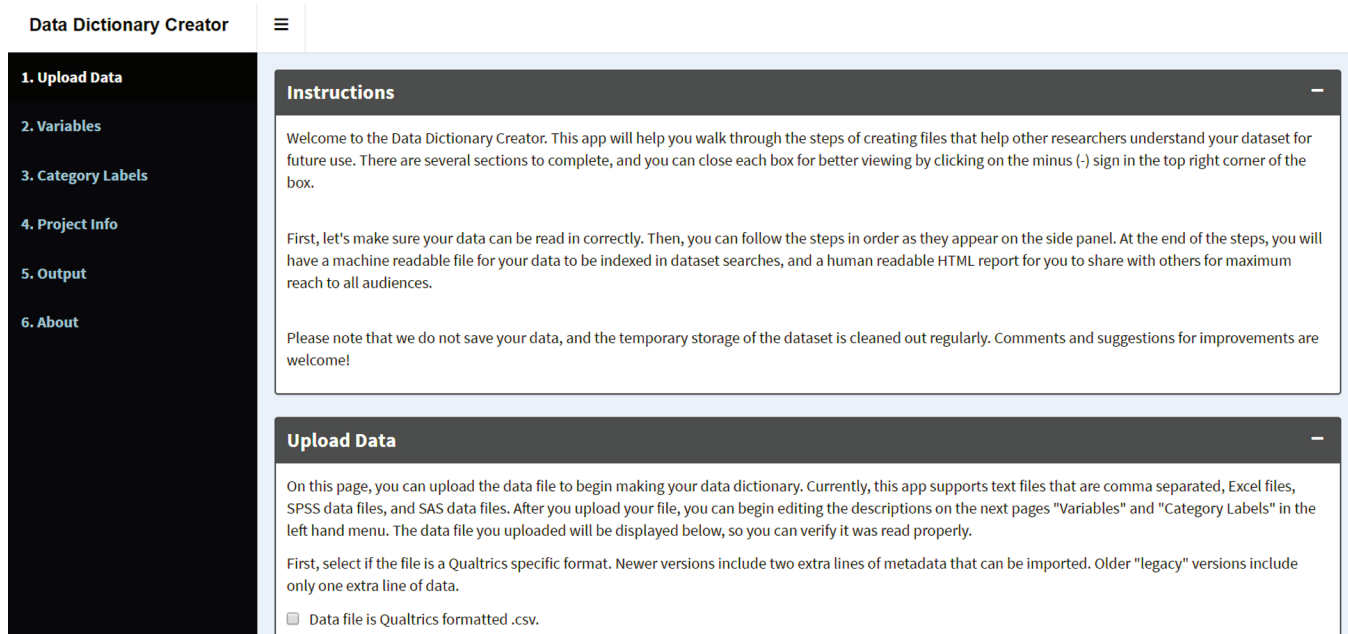


Fig. 5. Screenshot of the project-information interface for *Data Dictionary Creator*.

encourage R users to consult Arslan (2019) for a complete package tutorial. The Web interface for *codebook* 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 that the data can be indexed in Google Dataset Search; thus, the output is both human and machine readable. The data and codebook created using this Web application can be shared on sites such as 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.

Data Dictionary Creator

Data Dictionary Creator (Buchanan et al., 2019) breaks down metadata entry into five steps, as shown in the left side of Figure 5. First, the user uploads the data file for processing only (i.e., the data are not stored permanently). The uploaded data can be previewed to determine if they were imported correctly. The second step of the process involves entering the metadata for each column provided in the data set. The application automatically provides starting points for the number of unique values, missing values, variable type (e.g., character, numeric), and minimum and maximum values. A description of each column can be added, along with information about the levels or groups in the data and synonyms for the variables. Any embedded metadata from files such as SPSS, SAS, or Qualtrics csv files (e.g., some metadata are stored in the second row) are

imported into the description or category-label attributes for the third step. Category labels can be provided for both character and numeric data (e.g., responses on Likert-type scales that include labeled numbers), and these labels can quickly be copied over from one column to an entire scale. The fourth step is to enter overall project information, such as the citation, website, funders, dates of data collection, and authors. Finally, in the fifth step, users can download csv files of the metadata, a JSON-LD-formatted metadata file, and an RData file that includes the data set and descriptive information integrated together. This application is built with the *shiny R* package (Chang et al., 2019), and the default time-out options (i.e., the amount of time a user has to interact with an application) were increased to accommodate entering information for complex data sets.

Summary

In this Tutorial, we have detailed the concepts necessary to understand data dictionaries, codebooks, and metadata and provided information for researchers to create their own. This type of tutorial is especially critical as transparency practices become more commonplace and FAIR guidelines for sharing information and open data are implemented in journals. For example, the availability of large, open neuroimaging data sets led to the development of the BIDS, which defines standards for curating open neuroscientific data (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 such as *Nature Scientific*

Data, *Data in Brief*, and *Journal of Open Psychology Data*. This Tutorial and the accompanying videos on OSF provide a manageable first step toward generating understandable and reusable metadata for sharing and publication. The applications showcased here will continue to evolve as cohesive standards are formed through group discussion.

Transparency

Action Editor: Alex O. Holcombe

Editor: Daniel J. Simons

Author Contributions

E. M. Buchanan generated the idea for this manuscript, and S. E. Crain, A. L. Cunningham, H. R. Johnson, and H. Stash wrote the first draft. All the authors were involved in editing the original draft and making subsequent revisions. E. M. Buchanan programmed the *Data Dictionary Creator* application, receiving critical user feedback from S. E. Crain, A. L. Cunningham, H. R. Johnson, H. Stash, P. M. Isager, and R. Carlsson.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Open Practices

Open Data: not applicable

Open Materials: <https://osf.io/3y2ex/>


Preregistration: not applicable

All materials have been made publicly available via OSF and can be accessed at <https://osf.io/3y2ex/>. This article has received the badge for Open Materials. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iDs

Erin M. Buchanan  <https://orcid.org/0000-0002-9689-4189>

Peder Mortvedt Isager  <https://orcid.org/0000-0002-6922-3590>

Acknowledgments

We would like to thank the psych-DS team for their comments and contributions to the manuscript and *Data Dictionary Creator* application. Additionally, we thank Ruben Arslan and an anonymous reviewer for their helpful suggestions on improving this manuscript.

Prior Versions

A preprint of the submitted manuscript is available at <https://osf.io/vd4y3/>.

Notes

1. Note that *codebook* in italics refers to the application, not the document type.
2. You can view the data options at https://cran.r-project.org/web/packages/rio/vignettes/rio.html#supported_file_formats.

3. A participant is defined as a person or animal from a study or a unit from another type of studied data, such as stimuli or words.

References

- About Schema.org. (n.d.). <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>
- Buchanan, E. M., & Azevedo, F. (2019, July). *Statistics are useless without suitable data: How to implement and assess for data quality*. Workshop presented at the annual meeting of the Society for the Improvement of Psychological Science, Rotterdam, The Netherlands.
- Buchanan, E. M., DeBruine, L., & Mohr, A. H. (2019). *Data Dictionary Creator* (Version 0.1.1) [Computer software]. GitHub. <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 0.5.16) [Computer software]. Comprehensive R Archive Network. <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) [Computer software]. Comprehensive R Archive Network. <https://cran.r-project.org/package=shiny>
- Crockford, D. (n.d.). *Introducing JSON*. <https://www.json.org/json-en.html>
- Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., . . . Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3, Article 160044. <https://doi.org/10.1038/sdata.2016.44>
- Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsson, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Tessler, M. H., Lenne, R. L., Altman, S., Long, B., & 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), Article 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., Yap, M. J., Bengson, J. J., Niemeyer, D., & 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), Article e0218092. <https://doi.org/10.1371/journal.pone.0218092>

- Kline, M. (2018). *psych-DS*. GitHub. <https://github.com/psych-ds/psych-DS>
- McDaniel, G., & International Business Machines Corporation. (1994). *IBM dictionary of computing*. McGraw-Hill.
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B. A., Ram, K., Soderberg, C. K., Spies, J. R., Thaney, K., Updegrove, A., Woo, K. H., & Yarkoni, T. (2016). How open science helps researchers succeed. *eLife*, 5, Article e16800. <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*. Elsevier.
- Noy, N. (2018, September 5). Making it easier to discover datasets. *The Keyword*. <https://www.blog.google/products/search/making-it-easier-discover-datasets/>
- OSF Support. (n.d.). *How to make a data dictionary*. <http://help.osf.io/hc/en-us/articles/360019739054-How-to-Make-a-Data-Dictionary>
- Piwovar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, Article e175. <https://doi.org/10.7717/peerj.175>
- RDF Core Working Group. (2004). *RDF/XML syntax specification (revised)*. <https://www.w3.org/TR/REC-rdf-syntax/>
- Rolfs, 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. <https://doi.org/10.2139/ssrn.2694998>
- Smithsonian Libraries. (2018). *Describing your data: Data dictionaries*. <https://library.si.edu/sites/default/files/tutorial/pdf/datadictionaries20180226.pdf>
- Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). *Meaning in Life Questionnaire*. APA PsycTESTS. <https://doi.org/10.1037/t01074-000>
- University of Iowa Libraries. (n.d.). *Readme, data dictionaries, codebooks*. <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), Article 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)*. The Resilience Center.
- Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59(10). <https://doi.org/10.18637/jss.v059.i10>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, Article 160018. <https://doi.org/10.1038/sdata.2016.18>