

## 1. POR QUE PRECISAMOS DOCUMENTAR OS DADOS DE INVESTIGAÇÃO?

- A documentação dos dados de investigação é um passo crítico para determinar a sua usabilidade futura. Sem documentação apropriada, o conteúdo de uma base de dados pode ser ambíguo, ou mesmo inutilizável. Além disso, a documentação pode permitir a indexação dos dados por motores de busca aumentando o impacto da partilha dos dados
- Esta documentação é parte integrante dos **metadados**, ou seja, informação sobre os dados
- Os metadados permitem que os dados sejam facilmente localizados, compartilhados e usados de maneira eficiente e precisa

## 2. O QUE DEVE ESTAR CONTIDO NO CODEBOOK?

**Descrição da base de dados:** o que está apresentado em cada coluna, os valores válidos (min. max, missing), chave para códigos numéricos (e.g., 1 = feminino; 2 = masculino), tipo de informação que a variável representa (variável entre-sujeitos, e.g., idade, sexo, grupo; ou intra-sujeito, e.g., diferentes condições a que o participante foi exposto; diferentes medidas), instrumento usado (escala, tarefa padronizada, etc), unidade de medida (e.g., segundos, meses, anos; proporção de respostas corretas; escala, peso em kgs, altura em cm)

- Também pode conter informações sobre quando os dados foram recolhidos, responsável pela curadoria e data da última atualização
- Idealmente, nomeação das variáveis deve ser o mais transparente possível

## 3. COMO FAZER O ARQUIVO DE CODEBOOK?

- É imperativo incluir um codebook junto com a partilha dos dados
- O primeiro passo é definir qual será a audiência para qual o documento de metadados será produzido. O documento pode ser criado de forma a dar suporte ao usuário humano da base de dados, ou pode também ser criado de forma a permitir que motores de busca consigam encontrar os dados
- O arquivo de metadados pode ser criado em excel, word ou outro processador de texto. Estes arquivos são suficientes para a análise humana. Entretanto, também pode ser interessante utilizar ferramentas online que criam páginas de HTML e indicadores que permitem a indexação da base de dados por motores de busca. Estas ferramentas produzem dados que são compatíveis com linguagem de máquina. A seguir apresentamos duas ferramentas que podem ser úteis na criação destas documentações. A alternativa 1 cria documentação para o usuário humano. A alternativa 2 cria documento que permite a indexação por motores de busca

# PASSO-A-PASSO - ALTERNATIVA 1

1

Utilizar a ferramenta online disponível em:

<https://pmusfeld.shinyapps.io/codebook-generator/>

2

Visualização do website:

Utilize este botão para fazer o upload da base de dados

Se a base de dados contém nomes na primeira linha, deixe esta caixa marcada. Caso contrário remova

**Codebook Generator**

This is a simple app to generate a minimal codebook for your data set. To start, upload your data set in the panel below. By Clicking on *Start Codebook*, the app will read out the variable names of your data set and create a text input for each variable. Use the text input to describe the content of your variables. The app also computes some simple summary statistics of your variables. You can check and download your codebook at the end of the app.

**Note: The app was originally created for private purposes and is not tested properly. If you encounter any issues or want to make suggestions, please contact me.**  
Contact: Philipp Musfeld (philipp.musfeld@psychologie.uzh.ch)

The app currently supports any sort of text-files (e.g., txt, csv, tsv, dat), R-files (rda, rds and RData), SPSS-files (sav) and Excel-files (xls, xlsx)

Browse... No file selected

The first row of my file contains column labels

Start Codebook

Create your Codebook:

3

Poderá então entrar uma descrição de cada variável na base de dados.

Create your Codebook:

idCode:

Type of variable

Subject Identifier

Detailed variable description

This variable represents the anonymous code for the participants

languageGroup:

Type of variable

Between-subject variable

Detailed variable description

This variable indicates the language group of the participant (monolingual, bidialectal, bilingual)

4

Ao pé da página, visualize e faça o download do CODEBOOK:

Download and Print your Codebook:

Print Codebook

Download Codebook

Show 25 entries

Search:

variable	class	summary	missing_values	type	description
idCode	numeric	numeric with values between 1 and 111	0 missing values (NA)	Subject Identifier	This variable represents the anonymous code for the participants
languageGroup	factor	factor with 4 levels: Bidialectals, Bilingual-Dissimilar, Bilingual-Similar, Monolingual	0 missing values (NA)	Between-subject variable	This variable indicates the language group of the participant (monolingual, bidialectal, bilingual)
age	numeric	numeric with values between 18 and 33	0 missing values (NA)	Between-subject variable	The age of the participant in years
gender	numeric	numeric with values between 0 and 1	0 missing values (NA)	Between-subject variable	Sex of the participant (male = 1; female = 2)
Raven	numeric	numeric with values between 0 and 1	0 missing values (NA)	Dependent Variable	Proportion of correct responses in the short-version of the raven's matrice test
WM	numeric	numeric with values between 0 and 1	0 missing values (NA)	Dependent Variable	Proportion of correct responses in a working memory updating task with numerical values

Showing 1 to 6 of 6 entries

Previous 1 Next



# PASSO-A-PASSO - ALTERNATIVA 2

1

Utilizar a ferramenta online disponível em:

<https://opencpu.psych.bio.uni-goettingen.de/ocpu/library/codebook/www/>

2

Visualização do website:

Codebook generator .sav/.dta/.rds/other data file:  No file selected.

```
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 ---
```

Utilize este botão para fazer o upload da base de dados

Após o upload dos dados, utilize este botão para gerar o codebook

3

Customizar as seguintes linhas do código antes de criar o CODEBOOK:

Linha 02: "Codebook": Incluir nome mais específico (e.g. Codebook -Study XYZ)

Linha 28: "We collected the following data." - Substituir por uma descrição mais precisa

Linha 51: Remover símbolo de comentário (#) e substituir "My Awesome Dataset" por uma descrição mais apropriada

4

Clique no botão Generate codebook

Codebook generator .sav/.dta/.rds/other data file:  Results\_WM\_Reasoning\_Bilingualism.txt

```
1 ---
2 title: "Codebook - Cognitive effects of bidialectalism and bilingualism"
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 We collected the following data as part of the project XYZ funded by the Biasch Foundation.
28
29 ---{r codebook}
30 # omit the following lines, if your missing values are already properly labelled
31 codebook_data <- detect_missing(codebook_data,
32   only_labelled = TRUE, # only labelled values are autodetected as
33   # missing
34   negative_values_are_missing = FALSE, # negative values are missing values
35   ninety_nine_problems = TRUE, # 99/999 are missing values, if they
36   # are more than 5 MAD from the median
37 )
38
39 ---
```

Codebook - Cognitive effects of bidialectalism and bilingualism

We collected the following data as part of the project XYZ funded by the Biasch Foundation.

## No missing values.

Metadata

Description

Dataset name: Working Memory and Raven's performance of the monolingual, bidialectals, and bilingual groups

The dataset has N=99 rows and 6 columns. 99 rows have no missing values on any column.

► Metadata for search engines

Variables

5

Faça o download do CODEBOOK:

Um documento de formato HTML será salvo no computador. Adicionar no repositório juntamente com os dados. O documento conterá informações que permitirão a indexação dos dados por motores de pesquisa especializados na busca de dados (e.g., Google Dataset Search, <https://toolbox.google.com/datasetsearch>)