**Exercise: Creating metadata for a dataset**

1. Setup  
   We will use the *codebook* package to create a codebook for a fake dataset. Install codebook and some related packages in R by running the following code:  
   install.packages(c("codebook", "future", "ufs", "GGally", "GPArotation", "rio"))
2. Create a codebook file by running the code  
   codebook:: new\_codebook\_rmd()  
   This creates an Rmd file with a codebook template. We will adjust this template to create metadata for a fake dataset.
3. Download the fake dataset “fake\_data.RData” from Blackboard (in the folder for week 11). Put the file into the same folder as the codebook Rmd file you just created (for example, you can use the project folder of the Github repository from the previous exercise).
4. In the codebook Rmd, go to the second code chunk (```{r prepare\_codebook}).   
   Un-comment line 36 codebook\_data <- rio::import("mydata.sav") and change the file name to “fake\_data.RData”.
5. Knit the Rmd file and see what happens!  
   You should have created a codebook in the form of an html file. You should see an overview of each variable in the dataset.
6. This codebook is still very empty. Let’s add some metadata that would help search engines index this dataset and help other researchers understand what the dataset is about!  
   Go back to the Rmd file of the codebook.
   * Add metadata by using the following commands in the second code chunk:

Change the title of the codebook:  
metadata(codebook\_data)$name <- "[your title here]"  
metadata(codebook\_data)$description <- "[Describe the dataset in a few words, make it clear that this is fake data!]"  
metadata(codebook\_data)$creator <- list(

"@type" = "Person",

givenName = "Your first name",   
 familyName = "Your family name",

affiliation = list("@type" = "Organization",

name = "Utrecht University"))  
 metadata(codebook\_data)$datePublished <- "2022-11-16"

1. Add some variable labels: These are short descriptions of the variable. This is especially important when variable names are not self-explanatory. In our case, the variable names are in French, so it’s worth adding some labels. For example, you can label the transport variable with the description “mode of transportation” like this:  
   var\_label(codebook\_data$transport) <- "mode of transportation"
2. If you find the codebook in this default format too unwieldy, you can try making a more compact version. In the last chunk of the Rmd file, change  
   codebook(codebook\_data) to  
   compact\_codebook(codebook\_data)  
   Knit and see what the codebook looks like now!