



[DOI: 10.5281/zenodo.8433892](https://doi.org/10.5281/zenodo.8433892)

# How to prepare a publication package?

EUR Data Coffee Breaks

October 12 & 23, 2023

Eduard Klapwijk, PhD  
Research Data Steward, ESSB

**Erasmus University Rotterdam**



# The EUR Data Coffee Breaks

- Monthly meetings about Research Data Management by the EUR Data Stewards
- Short, focused, and informal
- 15-minute presentations
- Twice a month (hybrid + online)
- No recordings, slides available [here](#)



*Erasmus*

# Preserve and share data for research integrity

- *Netherlands Code of Conduct for Research Integrity* (2018)
- Transparency as an important principle of integrity
- Preserve and when possible, share data, code and research materials
  - *data is open and accessible to the extent possible and remains confidential to the extent necessary*



*Erasmus*

# Why a publication package?

- A way to make available data (or descriptions of data), materials and information to:
  - provide detailed insight in your methodology
  - provide the option to reproduce or replicate results (when applicable)
  - give others the opportunity to re-use it



Implementation of publication packages in this presentation is based on:  
DSW (Deans of Social Sciences in the Netherlands). (2022). Guideline for the archiving of academic research for Faculties of Behavioural and Social Sciences in the Netherlands (March 2022). Zenodo.  
<https://doi.org/10.5281/zenodo.7583831>

The Erasmus logo, featuring the word "Erasmus" in a stylized, cursive script.

# Archiving of academic research

## What to include in your publication package

Erasmus School of  
Social and  
Behavioural Sciences



### 1 Manuscript or publication

Must include a brief description of the problem definition, research design, data collection (sampling, selection and representativeness of informants) and methods used



### 2 Materials used

Include instructions, procedures, the design of the experiment and stimulus materials (interview guide, questionnaires, surveys, tests) necessary to replicate the research



### 3 Raw data files

Provide the most direct registration of behaviour or reactions of participants. Think of unfiltered export files of surveys, EEG measurements, recordings or transcripts. If needed, include all de-identification steps taken



### 4 Preprocessing computer code

Include code (such as Atlas.TI/SPSS/JASP syntax files, R scripts, etc.) describing the steps taken to process raw data into analysis data, including brief explanations of the steps in English



### 5 Processed data files

Provide data (either raw or processed) that were eventually analysed when preparing the article (e.g. a data file after transforming variables, after applying selection, etc.). If the raw data was analysed directly, step 3 suffices



### 6 Analysis computer code

Include code describing the steps taken to process the analysis data into the results reported in the manuscript, including brief explanations of the steps in English



### 7 Data management plan

Provide a copy of the most recent version of your data management plan



### 8 Readme file

Provide a clear readme describing who was involved in the project, when the data was collected, which documents and files can be found where and how to interpret them



### 9 Ethics documentation

Documents related to the ethical approval (e.g. approval letter, blank consent form)

Klapwijk, E.T., Tafese, E., & Schijven, B. (2023). Publication packages for academic research (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.7575566>

# Publication package example

- README.txt
- data
  - safi\_raw.csv
- data\_output
  - safi\_processed-for-plotting.csv
- documentation
  - dmp\_eur-pp\_v1.pdf
  - ethics\_approval\_letter.pdf
  - informed\_consent\_form.pdf
  - terms
    - eur\_ict\_policy\_20210713.pdf
    - eur\_rdm\_policy\_20200814.pdf
    - posit\_eula\_20221122.pdf
- manuscript\_rsos\_20230401.pdf
- materials
  - codebook\_variables\_safi.csv
  - survey\_questionnaire\_safi.csv
- scripts
  - analyses\_safi.html
  - analyses\_safi.qmd
  - preprocessing\_safi.html
  - preprocessing\_safi.qmd
  - references.bib
- yoda-metadata.json



May 22, 2023

Dataset Open Access

## EUR publication package example

Klapwijk, Eduard; Sunami, Nami; Mania, Joanna; Lushaj, Bora; Volkova, Anna

This publication package serves as an example for how a publication package could look like based on the [Guideline for the archiving of academic research for Faculties of Behavioural and Social Sciences in the Netherlands](#) (but might be useful for other disciplines and countries as well).

Dataset used in this example is the [SAFI Survey Results](#) dataset. The original dataset was generated by Philip Woodhouse, Gert Jan Veldwisch, Daniel Brockington, Hans C. Komakech, Angela Manjichi, Jean-Philippe Venot

Package is uploaded as ZIP file to preserve the folder structure. You can download the `eur-publication-package.zip` file and unzip it locally to explore the contents

Klapwijk, E.T., Sunami, N., Mania, J., Lushaj, B., & Volkova, A. (2023). EUR publication package example [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.795660>

# Manuscript or publication

- README.txt
- data
  - safi\_raw.csv
- data\_output
  - safi\_processed-for-plotting.csv
- documentation
  - dmp\_eur-pp\_v1.pdf
  - ethics\_approval\_letter.pdf
  - informed\_consent\_form.pdf
  - terms
    - eur\_ict\_policy\_20210713.pdf
    - eur\_rdm\_policy\_20200814.pdf
    - posit\_eula\_20221122.pdf
- manuscript\_rsos\_20230401.pdf
- materials
  - codebook\_variables\_safi.csv
  - survey\_questionnaire\_safi.csv
- scripts
  - analyses\_safi.html
  - analyses\_safi.qmd
  - preprocessing\_safi.html
  - preprocessing\_safi.qmd
  - references.bib
- yoda-metadata.json



## Manuscript or publication

Must include a brief description of the problem definition, research design, data collection (sampling, selection and representativeness of informants) and methods used

## ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research



Article submitted to journal

### Subject Areas:

social sciences, humanities, research data management

### Keywords:

archival package, publication package, SAFI dataset

### Author for correspondence:

EUR Data Stewards  
e-mail: [datasteward@eur.nl](mailto:datasteward@eur.nl)

## Manuscript using the SAFI dataset

Eduard Klapwijk<sup>1</sup>, Nami Sunami<sup>1</sup>, Joanna Mania<sup>1</sup>, Bora Lushaj<sup>1</sup>, Anna Volkova<sup>1</sup>

<sup>1</sup>Erasmus University Rotterdam, Rotterdam, Netherlands

This is a mock manuscript, produced to give an example of a complete archival package. One part of that package is the published (or accepted) manuscript or publication.

### 1. Introduction

This mock file is intended to serve as an example of a manuscript that should be part of the archival package. This manuscript should include a brief description of the problem definition, research design, data collection (sampling, selection and representativeness of informants) and methods used. An electronic version of the published manuscript will generally suffice. In this case we provide

# Materials used

- README.txt
- data
  - safi\_raw.csv
- data\_output
  - safi\_processed-for-plotting.csv
- documentation
  - dmp\_eur-pp\_v1.pdf
  - ethics\_approval\_letter.pdf
  - informed\_consent\_form.pdf
- terms
  - eur\_ict\_policy\_20210713.pdf
  - eur\_rdm\_policy\_20200814.pdf
  - posit\_eula\_20221122.pdf
- manuscript\_rsos\_20230401.pdf
- materials
  - codebook\_variables\_safi.csv
  - survey\_questionnaire\_safi.csv
- scripts
  - analyses\_safi.html
  - analyses\_safi.qmd
  - preprocessing\_safi.html
  - preprocessing\_safi.qmd
  - references.bib
- yoda-metadata.json



**Materials used**

Include instructions, procedures, the design of the experiment and stimulus materials (interview guide, questionnaires, surveys, tests) necessary to replicate the research

variable	unique_values	missing_values	levels	description	type	min
key_ID	129	0	NA	Unique Id for each observation	integer	1
village	3	0	Chirodzo, God, Ruaca	Village name	factor	
interview_date	19	0	NA	Date of interview	factor	2016-11
no_membrs	15	0	NA	Number of household members	integer	2
years_liv	50	0	NA	Number of years living in a given village	integer	1
respondent_wall_type	6	0	burntbricks, muddaub, burntbricks, cement, muddaub, sunbricks	Type of wall in the house	factor	
rooms	6	0	1, 2, 3, 4, 5, 8	Number of rooms for sleeping	integer	1
memb_assoc	3	0	no, NULL, yes	Membership irrigation association	factor	no
affect_conflicts	5	0	frequently, more_once, never, NULL, once	Frequency affected by a conflict	factor	frequent
liv_count	5	0	1, 2, 3, 4, 5	Livestock count	integer	1
items_owned	95	0	NA	Items owned by a household	factor	
no_meals	2	0	2, 3	Number of meals per day	integer	2
months_lack_food	31	0	NA	Month with no food available	factor	
instanceID	130	0	NA	Unique identifier for the form data submission	factor	

# Raw data files

README.txt

## data

safi\_raw.csv

## data\_output

safi\_processed-for-plotting.csv

## documentation

dmp\_eur-pp\_v1.pdf

ethics\_approval\_letter.pdf

informed\_consent\_form.pdf

## terms

eur\_ict\_policy\_20210713.pdf

eur\_rdm\_policy\_20200814.pdf

posit\_eula\_20221122.pdf

manuscript\_rsos\_20230401.pdf

## materials

codebook\_variables\_safi.csv

survey\_questionnaire\_safi.csv

## scripts

analyses\_safi.html

analyses\_safi.qmd

preprocessing\_safi.html

preprocessing\_safi.qmd

references.bib

yoda-metadata.json



**Raw data files**

Provide the most direct registration of behaviour or reactions of participants. Think of unfiltered export files of surveys, EEG measurements, recordings or transcripts. If needed, include all de-identification steps taken

key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	rooms	memb_assoc	affect_conflicts	liv_count	items_owned
1	God	2016-11-17T00:00:00Z	3	4	muddaub	1	NULL	NULL	1	bicycle;television;solar_panel;table
1	God	2016-11-17T00:00:00Z	7	9	muddaub	1	yes	once	3	cow_cart;bicycle;radio;cow_plough;solar_pa
3	God	2016-11-17T00:00:00Z	10	15	burntbricks	1	NULL	NULL	1	solar_torch
4	God	2016-11-17T00:00:00Z	7	6	burntbricks	1	NULL	NULL	2	bicycle;radio;cow_plough;solar_panel;mobile
5	God	2016-11-17T00:00:00Z	7	40	burntbricks	1	NULL	NULL	4	motorcyle;radio;cow_plough;mobile_phone
6	God	2016-11-17T00:00:00Z	3	3	muddaub	1	NULL	NULL	1	NULL
7	God	2016-11-17T00:00:00Z	6	38	muddaub	1	no	never	1	motorcyle;cow_plough
8	Chirodzo	2016-11-16T00:00:00Z	12	70	burntbricks	3	yes	never	2	motorcyle;bicycle;television;radio;cow_ploug
9	Chirodzo	2016-11-16T00:00:00Z	8	6	burntbricks	1	no	never	3	television;solar_panel;solar_torch
10	Chirodzo	2016-12-16T00:00:00Z	12	23	burntbricks	5	no	never	2	cow_cart;motorcyle;bicycle;television;radio;
11	God	2016-11-21T00:00:00Z	6	20	sunbricks	1	NULL	NULL	2	radio;cow_plough
12	God	2016-11-21T00:00:00Z	7	20	burntbricks	3	yes	never	2	cow_cart;bicycle;radio;cow_plough;table
13	God	2016-11-21T00:00:00Z	6	8	burntbricks	1	no	never	3	bicycle;radio;cow_plough;mobile_phone
14	God	2016-11-21T00:00:00Z	10	20	burntbricks	3	NULL	NULL	3	bicycle;radio;cow_plough;solar_panel;table;
15	God	2016-11-21T00:00:00Z	5	30	sunbricks	2	yes	once	3	bicycle;radio;cow_plough;solar_panel;table
16	God	2016-11-24T00:00:00Z	6	47	muddaub	1	NULL	NULL	4	radio;cow_plough;solar_panel;solar_torch
17	God	2016-11-21T00:00:00Z	8	20	sunbricks	1	NULL	NULL	1	mobile_phone

# Preprocessing computer code

- README.txt
- data
  - safi\_raw.csv
- data\_output
  - safi\_processed-for-plotting.csv
- documentation
  - dmp\_eur-pp\_v1.pdf
  - ethics\_approval\_letter.pdf
  - informed\_consent\_form.pdf
- terms
  - eur\_ict\_policy\_20210713.pdf
  - eur\_rdm\_policy\_20200814.pdf
  - posit\_eula\_20221122.pdf
- manuscript\_rsos\_20230401.pdf
- materials
  - codebook\_variables\_safi.csv
  - survey\_questionnaire\_safi.csv
- scripts
  - analyses\_safi.html
  - analyses\_safi.qmd
  - preprocessing\_safi.html
  - preprocessing\_safi.qmd
  - references.bib
- yoda-metadata.json

4



## Preprocessing computer code

Include code (such as Atlas.Ti/SPSS/JASP syntax files, R scripts, etc.) describing the steps taken to process raw data into analysis data, including brief explanations of the steps in English

## Preprocessing SAFI data

In this [Quarto](#) markdown document we share the computer code and describe the steps taken to process the raw data into analysis data used in the manuscript.

There are many ways to include computer code in your publication package, depending on the analysis tools you use. Using tools like [Quarto](#), [R Markdown](#), or [Jupyter](#) notebooks is a great way to share code and narrative text in one document.

### SAFI data

For this mock project we use (a cleaned version of) the SAFI dataset ([Woodhouse et al. 2018](#)), already in our `data` folder. The preprocessing is based on the Data Carpentry for Social Sciences and Humanities materials ([R for Social Scientists](#)).

For data wrangling we use two R packages that are part of the tidyverse: `dplyr` and `tidyr`. For pointing to the home location of our project we use the [here](#) package. For displaying the data load `kableExtra`.

```
library(tidyverse)
library(here)
library(kableExtra)
```

```
interviews_plotting <- interviews %>%
  # pivot wider by items_owned
  separate_rows(items_owned, sep = ";") %>%
  # if there were no items listed, changing NA to no_listed_items
  replace_na(list(items_owned = "no_listed_items")) %>%
  mutate(items_owned_logical = TRUE) %>%
  pivot_wider(names_from = items_owned,
             values_from = items_owned_logical,
             values_fill = list(items_owned_logical = FALSE)) %>%
  # pivot wider by months_lack_food
  separate_rows(months_lack_food, sep = ";") %>%
  mutate(months_lack_food_logical = TRUE) %>%
  pivot_wider(names_from = months_lack_food,
             values_from = months_lack_food_logical,
             values_fill = list(months_lack_food_logical = FALSE)) %>%
  # add some summary columns
  mutate(number_months_lack_food = rowSums(select(., Jan:May))) %>%
  mutate(number_items = rowSums(select(., bicycle:car)))
```

# Processed data files

- 📄 README.txt
- 📁 data
  - 📄 safi\_raw.csv
  - 📁 data\_output
    - 📄 safi\_processed-for-plotting.csv
- 📁 documentation
  - 📄 dmp\_eur-pp\_v1.pdf
  - 📄 ethics\_approval\_letter.pdf
  - 📄 informed\_consent\_form.pdf
- 📁 terms
  - 📄 eur\_ict\_policy\_20210713.pdf

- 📄 manuscript
- 📁 materials
  - 📄 code
  - 📄 surve
- 📁 scripts
  - 📄 analy
  - 📄 analy
  - 📄 prep
  - 📄 prep
  - 📄 refer
- 📄 yoda-metadata.json

key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	rooms	memb_assoc	affect_conflicts	liv_count	no_meals	instanceID	bicycle	television	solar_panel	table	cow_cart	radio	cow_plough	solar_torch	mo
1	God	2016-11-17T00:00:00Z	3	4	muddaub	1	NA	NA	1	2	uuid:ec241f2c-0609-46ed-b5e8-fe575f6cefef	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FAL
1	God	2016-11-17T00:00:00Z	7	9	muddaub	1	yes	once	3	2	uuid:099de9c9-3e5e-427b-8452-26250e840d6e	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRU
3	God	2016-11-17T00:00:00Z	10	15	burntbricks	1	NA	NA	1	2	uuid:193d7daf-9582-409b-bf09-027dd36f9007	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRU
4	God	2016-11-17T00:00:00Z	7	6	burntbricks	1	NA	NA	2	2	uuid:148d1105-778a-4755-aa71-281ead4a9773	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRU
5	God	2016-11-17T00:00:00Z	7	40	burntbricks	1	NA	NA	4	2	uuid:2c867811-9696-4966-9866-f35c3e97d02d	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRU
6	God	2016-11-17T00:00:00Z	3	3	muddaub	1	NA	NA	1	2	uuid:daa56c91-c8e3-44c3-a663-af6a49a2ca70	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FAL
7	God	2016-11-17T00:00:00Z	6	38	muddaub	1	no	never	1	3	uuid:ae20a58d-56f4-43d7-bafa-e7963d850844	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FAL
8	Chirodzo	2016-11-16T00:00:00Z	12	70	burntbricks	3	yes	never	2	2	uuid:d6cee930-7be1-4fd9-88c0-82a08f90fb5a	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FAL
9	Chirodzo	2016-11-16T00:00:00Z	8	6	burntbricks	1	no	never	3	3	uuid:846103d2-b1db-4055-b502-9cd510bb7b37	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FAL
10	Chirodzo	2016-12-16T00:00:00Z	12	23	burntbricks	5	no	never	2	3	uuid:8f4e49bc-da81-4356-ae34-e0d794a23721	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FAL
11	God	2016-11-21T00:00:00Z	6	20	sunbricks	1	NA	NA	2	2	uuid:d29b44e3-3348-4afc-aa4d-9eb34c89d483	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FAL
12	God	2016-11-21T00:00:00Z	7	20	burntbricks	3	yes	never	2	3	uuid:e6ee6269-b467-4e37-91fc-5e9ea934557	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FAL
13	God	2016-11-21T00:00:00Z	6	8	burntbricks	1	no	never	3	2	uuid:6c00c145-ee3b-409c-8c02-2c8d743b6918	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRU
14	God	2016-11-21T00:00:00Z	10	20	burntbricks	3	NA	NA	3	3	uuid:9b21467f-1116-4340-a3b1-1ab64f13c87d	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRU
15	God	2016-11-21T00:00:00Z	5	30	sunbricks	2	yes	once	3	2	uuid:a837e545-ff86-4a1c-a1a5-6186804b958f	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FAL
16	God	2016-11-24T00:00:00Z	6	47	muddaub	1	NA	NA	4	3	uuid:d17db52f-4b87-4768-b534-ea8f9704c565	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	FAL
17	God	2016-11-21T00:00:00Z	8	20	sunbricks	1	NA	NA	1	2	uuid:4707f3dc-df18-4348-9c2c-ee6c51e89b6b	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRU

5



## Processed data files

Provide data (either raw or processed) that were eventually analysed when preparing the article (e.g. a data file after transforming variables, after applying selection, etc.). If the raw data was analysed directly, step 3 suffices

# Analysis computer code

README.txt

## data

safi\_raw.csv

## data\_output

safi\_processed-for-plotting.csv

## documentation

dmp\_eur-pp\_v1.pdf

ethics\_approval\_letter.pdf

informed\_consent\_form.pdf

## terms

eur\_ict\_policy\_20210713.pdf

eur\_rdm\_policy\_20200814.pdf

posit\_eula\_20221122.pdf

manuscript\_rsos\_20230401.pdf

## materials

codebook\_variables\_safi.csv

survey\_questionnaire\_safi.csv

## scripts

analyses\_safi.html

analyses\_safi.qmd

preprocessing\_safi.html

preprocessing\_safi.qmd

references.bib

yoda-metadata.json

6



## Analysis computer code

Include code describing the steps taken to process the analysis data into the results reported in the manuscript, including brief explanations of the steps in English

## SAFI data analysis

In this [Quarto](#) markdown document we share the computer code and describe the steps taken to process the analysis data into the results reported in the manuscript.

There are many ways to include computer code in your publication package, depending on the analysis tools you use. Using tools like [Quarto](#), [R Markdown](#), or [Jupyter](#) notebooks is a great way to share code and narrative text in one document.

## Preparation

For the data analysis in this mock project we use the [safi\\_processed-for-plotting.csv](#) file that we prepared in the [preprocessing\\_safi](#) document.

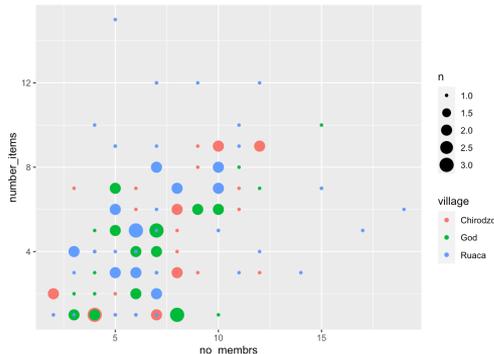
For data analysis and visualisation we use R packages that are part of the tidyverse: [ggplot2](#) and [dplyr](#). For pointing to the home location of our project we use the [here](#) package. For displaying the data we load [kableExtra](#).

```
library(tidyverse)
library(here)
library(kableExtra)
```

```
interviews <- read_csv(here("data/safi_raw.csv"), na = "NULL")
interviews_plotting <- read_csv(here("data_output/safi_processed-for-plotting.csv"))
```

## Plots

```
interviews_plotting %>%
  ggplot(aes(x = no_members, y = number_items, colour = village)) +
  geom_count()
```



# Data management plan

- 📄 README.txt
- 📁 data
  - 📄 safi\_raw.csv
- 📁 data\_output
  - 📄 safi\_processed-for-plotting.csv
- 📁 documentation
  - 📄 dmp\_eur-pp\_v1.pdf
  - 📄 ethics\_approval\_letter.pdf
  - 📄 informed\_consent\_form.pdf
- 📁 terms
  - 📄 eur\_ict\_policy\_20210713.pdf
  - 📄 eur\_rdm\_policy\_20200814.pdf
  - 📄 posit\_eula\_20221122.pdf
- 📄 manuscript\_rsos\_20230401.pdf
- 📁 materials
  - 📄 codebook\_variables\_safi.csv
  - 📄 survey\_questionnaire\_safi.csv
- 📁 scripts
  - 📄 analyses\_safi.html
  - 📄 analyses\_safi.qmd
  - 📄 preprocessing\_safi.html
  - 📄 preprocessing\_safi.qmd
  - 📄 references.bib
- 📄 yoda-metadata.json



**Data management plan**

Provide a copy of the most recent version of your data management plan

## EUR PUBLICATION PACKAGE

Please tick the following boxes if you agree to act according to the following terms:

- I will check and, if necessary, update my data management plan a minimum of once a year
- I will discuss the data management plan with my research team
- I will answer all questions truthfully and to the best of my knowledge

Support in writing a data management plan is available through the [faculty Data Stewards](#). Which research support professional is available for you?

- Data Steward of my own faculty - ESHCC

Scientific research must be conducted in line with existing guidelines on good research practices and integrity. Please tick the boxes if you have read and understand these guidelines and will act accordingly.

- The Netherlands Code of Conduct for Research Integrity (VSNU, 2018)
- The European Code of Conduct for Research Integrity (ALLEA, 2017)

ADMINISTRATION & PROJECT DESCRIPTION



# Readme file

README.txt

data

- safi\_raw.csv

data\_output

- safi\_processed-for-plotting.csv

documentation

- dmp\_eur-pp\_v1.pdf
- ethics\_approval\_letter.pdf
- informed\_consent\_form.pdf
- terms

- eur\_ict\_policy\_20210713
- eur\_rdm\_policy\_2020081
- posit\_eula\_20221122.pdf

manuscript\_rsos\_20230401.pdf

materials

- codebook\_variables\_safi.csv
- survey\_questionnaire\_safi.csv

scripts

- analyses\_safi.html
- analyses\_safi.qmd
- preprocessing\_safi.html
- preprocessing\_safi.qmd
- references.bib

yoda-metadata.json



**Readme file**

Provide a clear readme describing who was involved in the project, when the data was collected, which documents and files can be found where and how to interpret them

```
README.txt
This is a mock readme file that was created on 2023-04-12 by EUR Research Data Stewards.

It serves as an example for how a publication package could look like based on the Guideline (see
https://doi.org/10.5281/zenodo.7583831) for the archiving of academic research for Faculties of
Behavioural and Social Sciences in the Netherlands (but might be useful for other disciplines and
countries as well).
Dataset used in this example is the SAFI Survey Results: https://doi.org/10.6084/
m9.figshare.6262019.v4
The original dataset was generated by Philip Woodhouse, Gert Jan Veldwisch, Daniel Brockington, Hans
C. Komakech, Angela Manjichi, Jean-Philippe Venot

GENERAL INFORMATION

Title of the project: EUR Publication Package

Principal Investigator Information
Name: Eduard Klapwijk
ORCID: 0000-0002-8936-0365
Institution: Erasmus University Rotterdam
Email: datasteward@essb.eur.nl

Co-investigator Information
Name: Nami Sunami
ORCID: 0000-0001-5482-8370
Institution: Erasmus University Rotterdam
Email: data@erim.eur.nl
```



# Ethics documentation

- README.txt
- data
  - safi\_raw.csv
- data\_output
  - safi\_processed-for-plotting.csv
- documentation
  - dmp\_eur\_pp\_v1.pdf
  - ethics\_approval\_letter.pdf
  - informed\_consent\_form.pdf
- terms
  - eur\_ict\_policy\_202107
  - eur\_rdm\_policy\_20200
  - posit\_eula\_20221122.p
- manuscript\_rsos\_20230401.pdf
- materials
  - codebook\_variables\_safi.csv
  - survey\_questionnaire\_safi.csv
- scripts
  - analyses\_safi.html
  - analyses\_safi.qmd
  - preprocessing\_safi.html
  - preprocessing\_safi.qmd
  - references.bib
- yoda-metadata.json



9

## Ethics documentation

Documents related to the ethical approval (e.g. approval letter, blank consent form)

Erasmus  
University  
Rotterdam

Date  
XXth Telona, 2328BY

Subject  
Approval ethics review application, "EUR  
Publication Package"

Our reference  
ETH-123456789

Your reference

Page  
1/1

Department  
Department of Mock Project

Visiting address  
Erasmus School of JEDI  
Burgemeester Oudlaan 50  
Jedi Building 1138

Postal address  
P.O. Box 1738  
3000 DR Rotterdam  
The Netherlands

T +31 10 408 1138X  
E ethics-at-yoda@eur.nl  
W www.eur.nl/English

Dear EUR Data Stewards,

On the XXth Telona, 2328BY, the Internal Review Board of the Erasmus School of JEDI received your request for approval of your research project "EUR Publication Package". Thank you for submitting your research proposal.

Hereby, I am pleased to inform you that the Board has decided to advise positively on your application. Based on the information you have provided, we conclude that your research proposal sufficiently respects the safety and rights of participants and recognises the responsibilities of the researcher(s) involved and of their host and sponsoring organisations. The approval decision is valid for three years.

Once you start your research, please be aware that you are responsible for adhering to the Netherlands Code of Conduct for Research Integrity whether working individually or in a team. Please note that you are required to submit an amendment in case any significant changes occur.

Yours sincerely,

On behalf of the Rotterdam School of JEDI, IRB

## Information and consent form

### Mock Project using the SAFI Dataset

#### Introduction

We are research data stewards and I work for Erasmus University Rotterdam. I am conducting research on how we can use the Yoda Archive, the place to store materials about research.

I will explain the study below. If you have any questions, please ask me. While reading, you can mark parts of the text that are unclear to you.

If you want to participate in the study, you can indicate this at the end of this form.

#### What is the research about?

The research is about whether we can make a mock project

#### Why are we asking you to participate?

We ask you to participate because your experience as a citizen helps us to understand the world.

#### What can you expect?

The study lasts 30 minutes.

If you participate in this study, you will take part in:

#### A SURVEY

You will receive a questionnaire by email or post.  
You can fill in the questionnaire yourself or together with me.

#### You decide whether to participate

Participation in this study is completely voluntary. [Not participating will not affect your work or work-related assessments or reports/ if you do not participate then your [type of care/ treatment] will continue as usual.]

# Publication packages for qualitative data

- For qualitative, interpretative methodologies:
  - Transparency possible, reproduction harder
  - Data should be preserved, only documented exceptions possible
- Additional information to record when doing fieldwork:
  - Dates of data collection
  - Type of activities (e.g., observations, conversations)
  - Interview and observation guides (if available)
  - Hard evidence of the period spent in the field



*Erasmus*

# What to share and what to archive?



## Data Repository:

- Data package with curated data
- Used materials, e.g. protocols, informed consent form (NOT signed)
- Usually NO personal data
- ...
- README file



## Archive:

- Raw data AND processed data
- Ethical review approval
- Informed consent forms (signed)
- Any other project documents (agreements, etc.)
- ...
- README file

*Erasmus*

# Where to share?



- Discipline-specific repositories (overview at [re3data](#))
  - preferably trusted repositories (e.g., CoreTrustSeal, ISO certified)
  - at least using persistent identifiers such as [DOI](#)
- Institutional repository
  - at EUR: the [EUR Data Repository](#)
- Generic data repository
  - e.g., [Zenodo](#), [Open Science Framework](#), [DANS Easy](#)





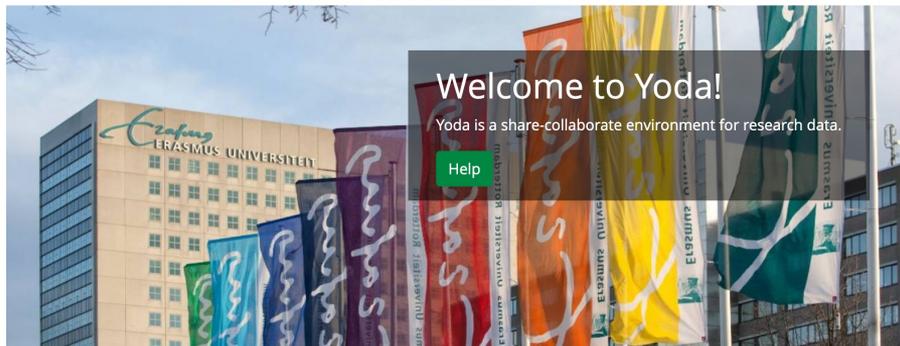
# Where to archive?

- Digital archive for long-term preservation at EUR
  - Based on EUR Yoda Vault, access via Yoda Portal
  - Secure storage for >10 years



Sign in

Yoda Portal



# Some practical takeaways

- Start planning early on 
- Prepare your data for preservation and re-use (e.g., use open file formats) 
- Document your data, keep it organized 
- Richly describe your data 

*Erasmus*

# Questions



*Erasmus*

# Thank you!

Coming up:

November - Privacy by design

December - Sharing qualitative data

January - Do's and Don'ts for data management plans

See [schedule](#) for more information

Do you have suggestion for a topic?

Contact us via [EUR Research Data Stewards](#)

