

# Data Management Workshop

26 March 2021

Agenda (in CET):

14:00	Intro Data Management
14:15	Icebreaker / Reflections assignments
14:35	Break out room 1
14:50	Reflection break out room 1
15:00	Break
15:15	Break out room 2
15:30	Reflection break out room 2
15:40	Intro Open Data
15:50	Wrap up, reflection workshop + feedback
16:00	The end

## Code of Conduct

- Please be inclusive and constructive
- Please be patient with the technology/each other/yourself

## Icebreaker

**Name / pronouns / favourite cake/pie/snack?**

- Esther Plomp / she,her / cashewnut pie

Briney et al. 2020

**Any questions/comments regarding this article? Was it helpful?**

- ...

# README template

**Was it difficult to set up a template? Did you find it helpful or not?**

- •  
•  
•  
•  
•  
•  
•  
•  
•

- 
- 
- 
- 
- 
- 
- 
- 

## Breakout room 1

1. Say hi and discuss one thing from this week that you learned or enjoyed
2. Share your README templates with each other through screen share or by reading part of the text to the person(s) you're grouped with or share the text in the Zoomchat:
  - Are you able to understand everything in the README file?
  - Is all of the information helpful?
  - Would you be able to (re)use the dataset without communicating with the creator of the data/README file: if not, what information are you missing?
  - What is the most important information about a dataset you would need in order to work with it?

### Optional notes:

#### Room 1:

- 
- 
- 

#### Room 2:

- 
- 
- 

#### Room 3:

-

- 
- 

**Room 4:**

- 
- 
- 

**Room 5:**

- 
- 
- 

**Room 5:**

- 
- 
- 

**Room 6:**

- 
- 
- 

**Room 7:**

- 
- 
-

## Breakout room 2

1. Say hi and discuss one thing from this week that you learned or enjoyed
2. Share your README templates with each other through screen share or by reading part of the text to the person(s) you're grouped with or share the text in the Zoomchat:
  - Are you able to understand everything in the README file?
  - Is all of the information helpful?
  - Would you be able to (re)use the dataset without communicating with the creator of the data/README file: if not, what information are you missing?
  - What is the most important information about a dataset you would need in order to work with it?

### Optional notes:

#### Room 1:

- 
- 
- 

#### Room 2:

- 
- 
- 

#### Room 3:

- 
- 
- 

#### Room 4:

- 
- 
- 

#### Room 5:

- 
- 
- 

#### Room 5:

-

- 
- 

**Room 6:**

- 
- 
- 

**Room 7:**

- 
- 
-

## Reflection

**Was there anything new you've learned? Anything you can apply to your research?**

- • • • •



## Feedback

## What went well?

- • • • •

## What could be improved?

- • • • •

## Resources and further help:

- Check out your local Open Science Community ([Groningen](#))
- Read [The Turing Way](#) to learn more about reproducible data science (and contribute if you see anything that is missing!)
- Get in touch with your local Research Data Management support from the Library
  - [Groningen](#)
  - [Cambridge](#)
  - [York](#)
- Check out the [Open Research Calendar](#) for updates about Open Science events
- Get started with programming: check out the [Carpentries](#) or our local graduate school for any workshops that are organised. If you already know a bit of coding, try the materials of the [Code Refinery](#).
- Ben Marwick has [set up a list](#) of relevant R packages for archaeologists, as well as a list with publications that have shared their code. Have a look to see how others have shared their data and code!
- For more information about R packages to make your code more reproducible, see [rOpenSci](#). For example, [rtools](#) is a package that helps you to improve the reproducibility of your research.
- There's an [Rchaeology Twitter](#)/Slack
- *"A Standard for the Scholarly Citation of Archaeological Data as an Incentive to Data Sharing"* by Ben Marwick and Suzanne Pilaar Birch, <https://doi.org/10.1017/aap.2018.3>
- *"Opening up the casket: why and how to share your research data"* by Esther Plomp, <https://doi.org/10.5281/zenodo.3543769>
- The [Center for Open Science](#) offers free storage and functions like a repository. It also has a lot of information on Open Science initiatives.
- You can work collaboratively on code using platforms such as [GitHub](#) / [GitLab](#)
- [Zenodo](#) is a good repository to share data/code/presentations and other materials and receive credits for them! See also [EASY](#), [Archaeology Data Services](#), [Open Context](#) & [Qualitative Data Repository](#).
- Check out the [RIOT Science Club](#) for open research events and information
- For a good read about some of the problems that Open Science tries to address, check out the book *Science Fictions* by Stuart Ritchie.
- Join a [ReproducibiliTea](#) club to discuss papers
  - [Groningen](#)

- [UiO](#)

## R

- [R data science and visualisation videos](#) by Danielle Navarro (beginner)
- [Slides from R seminars](#) by Danielle Navarro (beginner)
- [Learning Statistics with R](#) by Danielle Navarro
- [Data visualisation](#) in R by Danielle Navarro (intermediate)
- [R for data science](#)
- [Intro to R](#) by Andrew Jones and Ozayr Mohammed
- [Writing dynamic and reproducible documents - and introduction to R-markdown](#) by Olivier Gimenez ([GitHub](#))
- [R Markdown](#) by Danielle Navarro (beginner)
- [Reproducible Science with R](#) and Rstudio by Olivier Gimenez
- [R for beginners](#) by Ihnwhi Heo, Duco Veen, and Rens van de Schoot
- [ggplot2](#) by Hadley Wickham, Danielle Navarro, and Thomas Lin Pedersen
- ggplot2 workshop [part 1](#) and [2](#) by Thomas Lin Pedersen (YouTube)
- [ggplot 2 tutorial for beginners](#)
- [A ggplot2 tutorial for beautiful plotting in R](#) by Cedric Scherer
- [Creating and using custom ggplot2 themes](#) by Thomas Mock
- [ggplot 2 scales](#)
- [Multiple ggplot2 plots](#)
- [Fundamentals of Data Visualization](#) by Claus O. Wilke
- [Data Visualization, A practical introduction](#) by Kieran Healy
- [The Data Validation Cookbook](#) by Mark P.J. van der Loo
- [What They Forgot to Teach You About R](#) by Jennifer Bryan and Jim Hester
- [Advanced R](#) by Hadley Wickham
- [Rmarkdown examples/blogs/tutorials](#)
- [RMarkdown Driven Development](#) (Intro to RMarkdown, projects, packages)
- [Dive into dplyr](#)
- [dplyr tutorial](#) by Allison Horst (beginner)
- [Missing values](#) by Allison Horst (beginner) ([see video here](#); [slides here](#))
- [25 days of chRistmas](#) by Kiirsti Owen, with [videos by Jason Winget](#) (very beginner)

- [R for Excel users](#) by Julie Lowndes & Allison Horst
- [Reproducible Research Data & Project Management in R](#) by Anna Krystalli
- [Reproducible Analyses in R](#) (beginner)
- [Improve your workflow for reproducible science](#) (intermediate)
- [Happy Git and GitHub for the useR](#)
  - See [here](#) for a shorter version
- [WORCS, Workflow for Open Reproducible Code in Science](#)
- [Renv package explained](#) by Kevin Ushey
- [Structuring R projects](#)
- [Reproducible Research with R](#) (research compendia and rrttools, Holepunch and Binder, ReproHack)
- Package to help you share code to ask questions: [reprex](#)
- [Reproducible research with workflowr](#)
- [Exploring missing values in naniar](#) by Allison Horst

#### Open Refine

- [OpenRefine user manual](#)
- [Tutorial](#) by the Programming Historian
- [Tutorial](#) by Digitalnomad
- Introduction to [Digital Humanities](#) with Open Refine