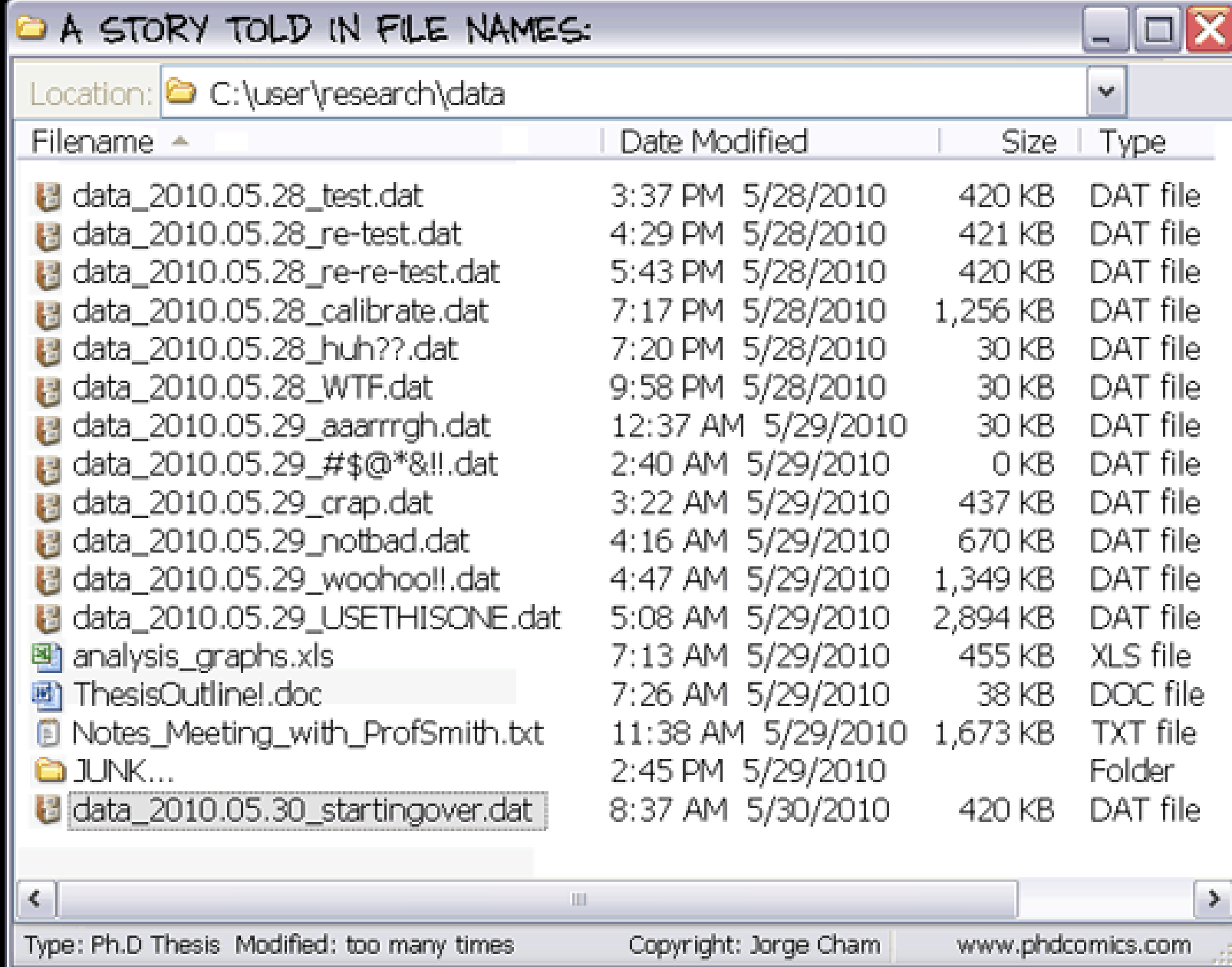


# Research Project management 101

CHRIS HJ HARTGERINK, JANUARY 22 2016





# "FINAL".doc



FINAL.doc!



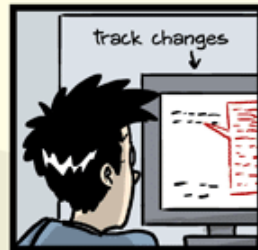
FINAL\_rev.2.doc



FINAL\_rev.6.COMMENTS.doc



FINAL\_rev.8.comments5.  
CORRECTIONS.doc



FINAL\_rev.18.comments7.  
corrections9.MORE.30.doc



FINAL\_rev.22.comments49.  
corrections.10.#@\$%WHYDID  
ICOMETOGRADSCHOOL????.doc





# DATA PACKAGE AUDIT







Which data  
selection?

Which  
datafile?

Which  
analyses?

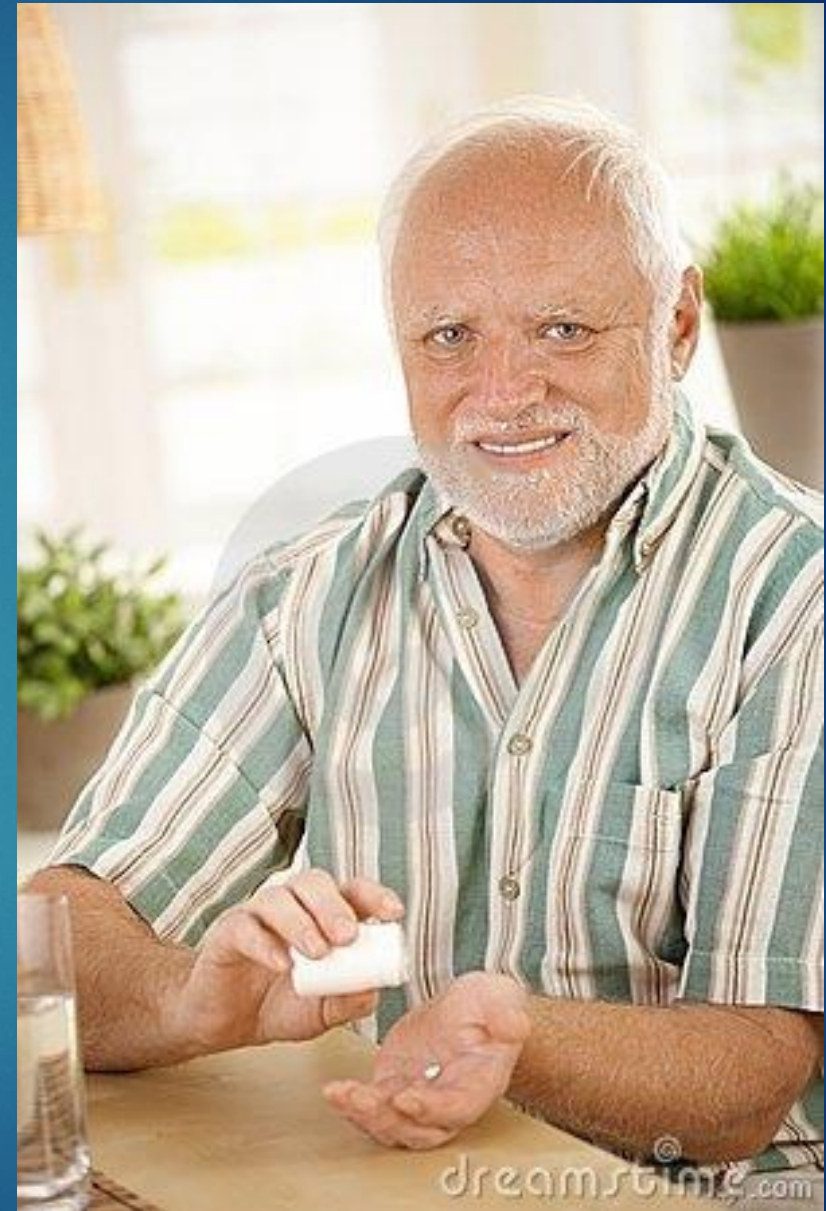


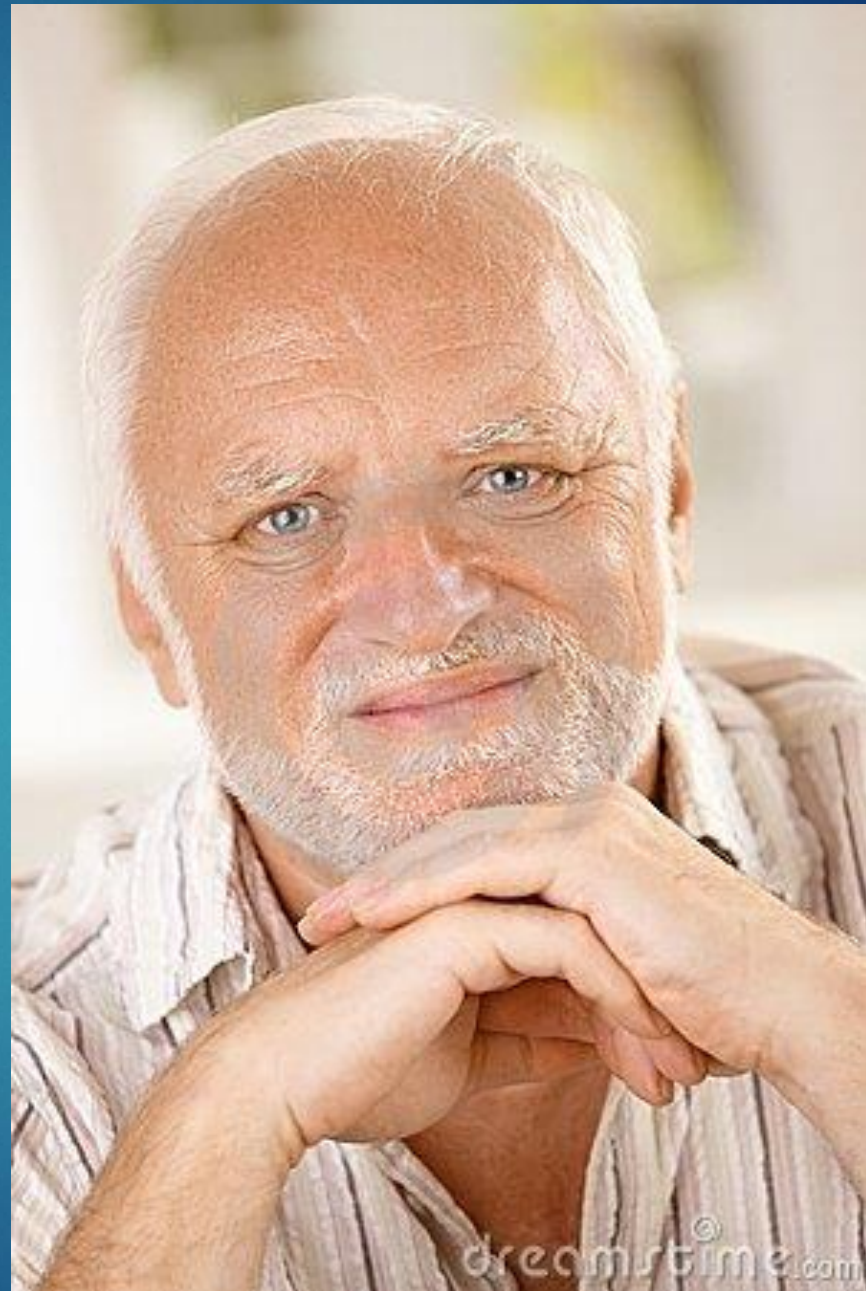
Which  
secondary  
analyses?

Which  
covariates?

Which variable  
transformations?



















# DATA REQUEST







Don't be like Harold

## **VSNU:**

*Information presented must be verifiable. If research results are published it must be clear what the data and conclusions are based upon, where they have been taken from and where they can be verified.*

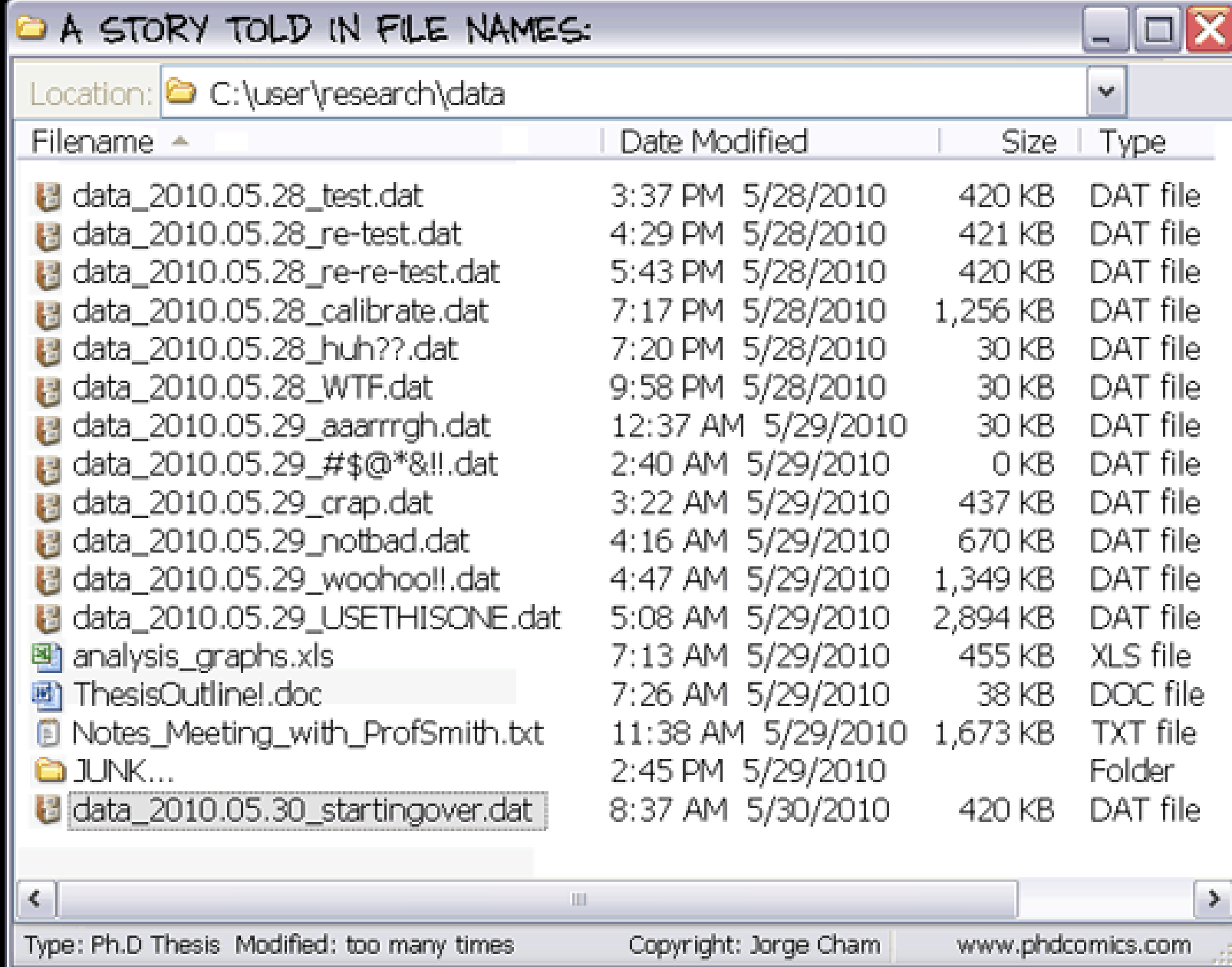


**Ask ourselves continuously:**

*Can my future-self or anyone else  
reproduce this result within a reasonable  
amount of time?*

# Ensure that project is...

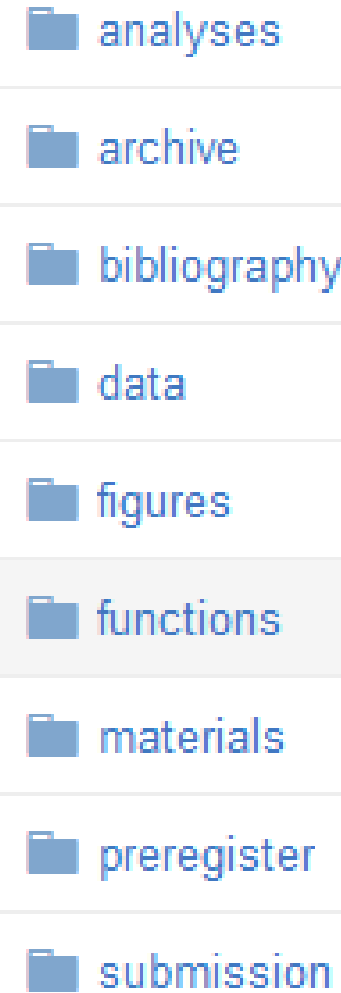
- ▶ Structured
- ▶ Logged changes
- ▶ Reproducible
- ▶ Shared
- ▶ (Continuously) backed-up
- ▶ Future proof





# Structuring the project

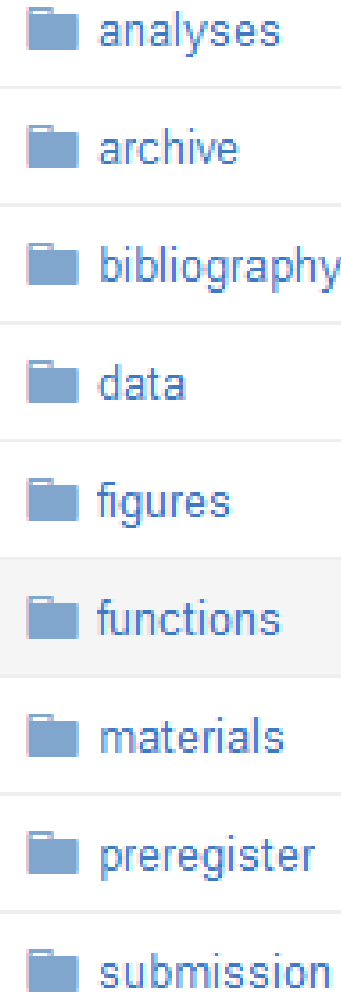
- ▶ Different folders for different aspects of your research



- analyses
- archive
- bibliography
- data
- figures
- functions
- materials
- preregister
- submission

# Structuring the project

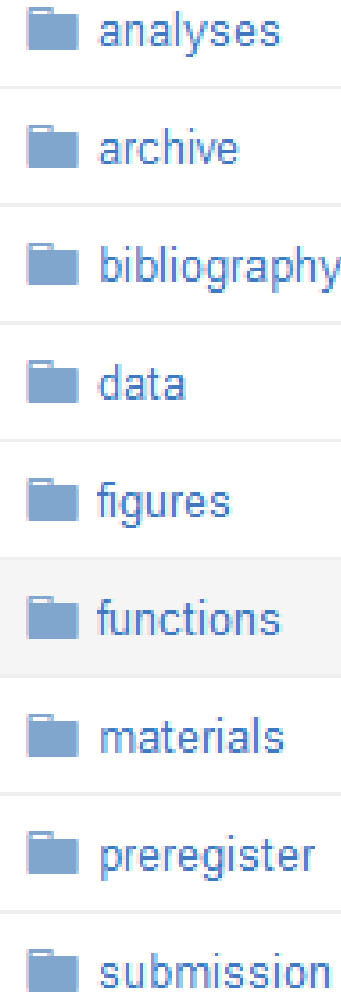
- ▶ Different folders for different aspects of your research
- ▶ Minimize folder levels
  - ▶ data/study 1/pilot study/ is less clear than data/pilot study 1



- analyses
- archive
- bibliography
- data
- figures
- functions
- materials
- preregister
- submission

# Structuring the project

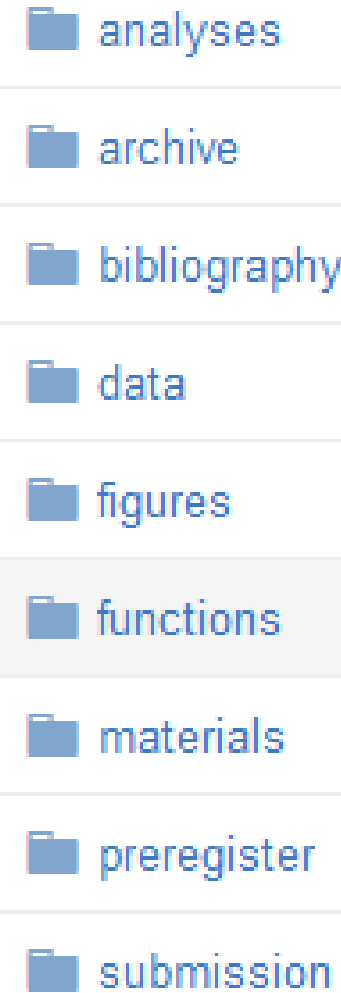
- ▶ Different folders for different aspects of your research
- ▶ Minimize folder levels
  - ▶ data/study 1/pilot study/ is less clear than data/pilot study 1
  - ▶ Try to prevent going three levels deep



- analyses
- archive
- bibliography
- data
- figures
- functions
- materials
- preregister
- submission

# Structuring the project

► Datestamp time sensitive files



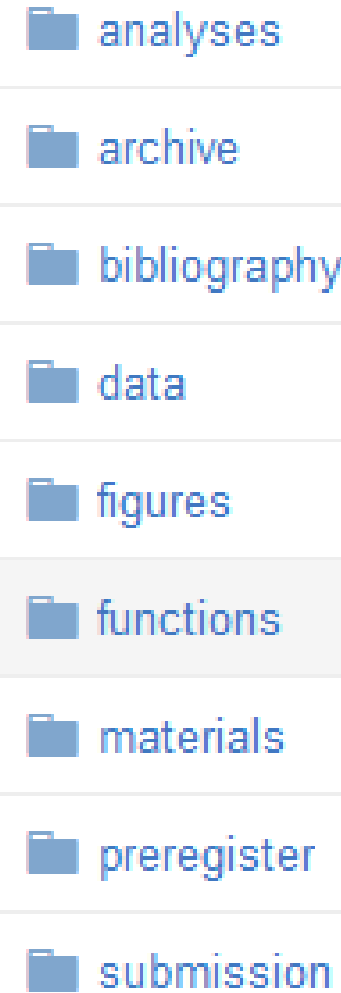
A vertical list of project folders, each preceded by a blue folder icon. The folders are: analyses, archive, bibliography, data, figures, functions, materials, preregister, and submission. The 'functions' folder is highlighted with a light gray background.

- analyses
- archive
- bibliography
- data
- figures
- functions
- materials
- preregister
- submission



# Structuring the project

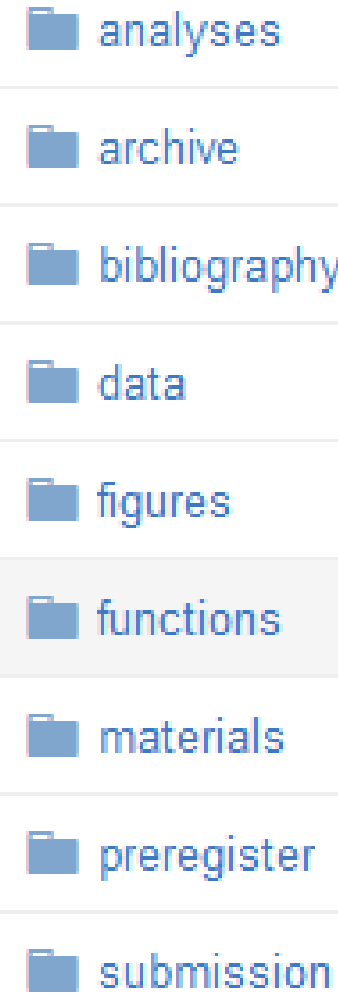
- ▶ Datestamp time sensitive files
  - ▶ Continuous data collection?



- analyses
- archive
- bibliography
- data
- figures
- functions
- materials
- preregister
- submission

# Structuring the project

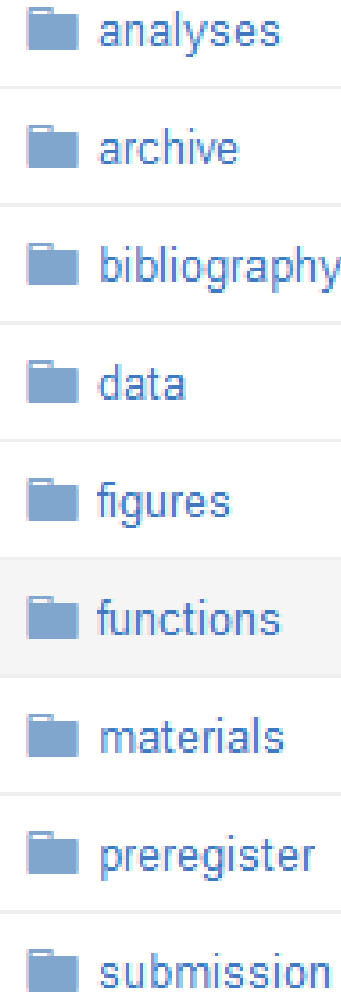
- ▶ Datestamp time sensitive files
  - ▶ Continuous data collection?
  - ▶ Emails/letters?



- analyses
- archive
- bibliography
- data
- figures
- functions
- materials
- preregister
- submission

# Structuring the project

- ▶ Datestamp time sensitive files
  - ▶ Continuous data collection?
  - ▶ Emails/letters?
  - ▶ Want to be able to chronologically sort something?

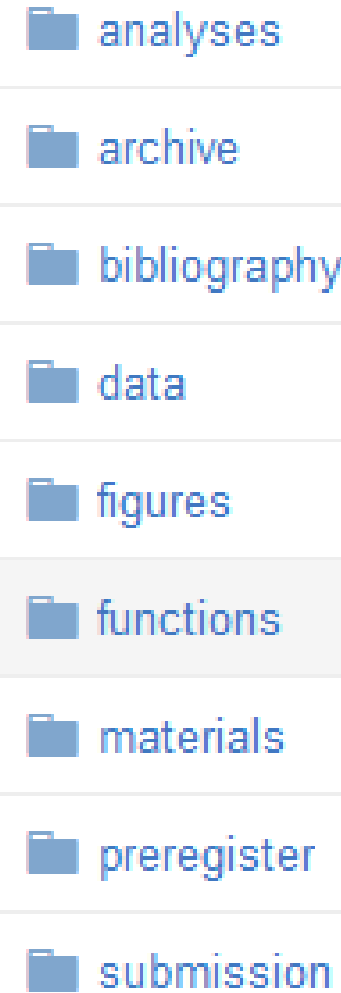


- analyses
- archive
- bibliography
- data
- figures
- functions
- materials
- preregister
- submission



# Structuring the project

- ▶ Datestamp time sensitive files
  - ▶ Continuous data collection?
  - ▶ Emails/letters?
  - ▶ Want to be able to chronologically sort something?
  - ▶ 22 January 2016 →  
20160122 filename.pdf



- analyses
- archive
- bibliography
- data
- figures
- functions
- materials
- preregister
- submission

Would you feel confident  
giving your project folder to  
a co-author?

# Logging changes

- ▶ Some files obviously change
  - ▶ Syntax files
  - ▶ Manuscript



# Logging changes

- ▶ Some files obviously change
  - ▶ Syntax files
  - ▶ Manuscript
- ▶ Some files obviously should not change
  - ▶ Raw data

# Logging changes

- ▶ Some files obviously change
  - ▶ Syntax files
  - ▶ Manuscript
- ▶ Some files obviously should not change
  - ▶ Raw data
- ▶ How to know if changes are made, what the changes are, and who made the changes?

# Version control



Version control  
=  
track changes for files

All times displayed at +0100 UTC offset.

- 2016-01-05 11:16 AM** CHJ Hartgerink updated file [analysis/master.R](#) to GitHub repo chartgerink / 2014bump in [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-12-20 10:37 AM** CHJ Hartgerink updated file [analysis/master.R](#) to GitHub repo chartgerink / 2014bump in [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-12-18 05:33 PM** CHJ Hartgerink updated file [analysis/master.R](#) to GitHub repo chartgerink / 2014bump in [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-11-30 11:51 AM** CHJH updated file [analysis/master.R](#) to GitHub repo chartgerink / 2014bump in [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-11-30 11:51 AM** CHJH added file [desktop.ini](#) to GitHub repo chartgerink / 2014bump to [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-11-16 10:00 AM** C.H.J. Hartgerink updated the license of [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-10-29 09:50 AM** CHJ Hartgerink removed file [submission/distributions-p\\_JMW mva.docx](#) from GitHub repo chartgerink / 2014bump from [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-10-29 08:47 AM** CHJ Hartgerink added file [submission/distributions-p\\_JMW mva.docx](#) to GitHub repo chartgerink / 2014bump to [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-09-21 03:46 PM** CHJ Hartgerink added file [archive/library.bib](#) to GitHub repo chartgerink / 2014bump to [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-09-21 09:01 AM** CHJ Hartgerink removed file [Info peek statistic.docx](#) from GitHub repo chartgerink / 2014bump from [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)

All times displayed at +0100 UTC offset.

- 2016-01-05 11:16 AM** CHJ Hartgerink updated file [analysis/master.R](#) to GitHub repo chartgerink / 2014bump in [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-12-20 10:37 AM** CHJ Hartgerink updated file [analysis/master.R](#) to GitHub repo chartgerink / 2014bump in [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-12-18 05:33 PM** CHJ Hartgerink updated file [analysis/master.R](#) to GitHub repo chartgerink / 2014bump in [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-11-30 11:51 AM** CHJH updated file [analysis/master.R](#) to GitHub repo chartgerink / 2014bump in [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-11-30 11:51 AM** CHJH added file [desktop.ini](#) to GitHub repo chartgerink / 2014bump to [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-11-16 10:00 AM** C.H.J. Hartgerink updated the license of [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-10-29 09:50 AM** CHJ Hartgerink removed file [submission/distributions-p\\_JMW mva.docx](#) from GitHub repo chartgerink / 2014bump from [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-10-29 08:47 AM** CHJ Hartgerink added file [submission/distributions-p\\_JMW mva.docx](#) to GitHub repo chartgerink / 2014bump to [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-09-21 03:46 PM** CHJ Hartgerink added file [archive/library.bib](#) to GitHub repo chartgerink / 2014bump to [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-09-21 09:01 AM** CHJ Hartgerink removed file [Info peek statistic.docx](#) from GitHub repo chartgerink / 2014bump from [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)



All times displayed at +0100 UTC offset.

**2016-01-05 11:16 AM**

CHJ Hartgerink updated file `analysis/master.R` to GitHub repo chartgerink / 2014bump in **Distributions of p-values smaller than .05 in Psychology: What is going on?**

**2015-12-20 10:37 AM**

CHJ Hartgerink updated file `analysis/master.R` to GitHub repo chartgerink / 2014bump in **Distributions of p-values smaller than .05 in Psychology: What is going on?**

**2015-12-18 05:33 PM**

CHJ Hartgerink updated file `analysis/master.R` to GitHub repo chartgerink / 2014bump in **Distributions of p-values smaller than .05 in Psychology: What is going on?**

**2015-11-30 11:51 AM**

CHJH updated file `analysis/master.R` to GitHub repo chartgerink / 2014bump in **Distributions of p-values smaller than .05 in Psychology: What is going on?**

**2015-11-30 11:51 AM**

CHJH added file `desktop.ini` to GitHub repo chartgerink / 2014bump to **Distributions of p-values smaller than .05 in Psychology: What is going on?**

**2015-11-16 10:00 AM**

C.H.J. Hartgerink updated the license of **Distributions of p-values smaller than .05 in Psychology: What is going on?**

**2015-10-29 09:50 AM**

CHJ Hartgerink removed file `submission/distributions-p_JMW mva.docx` from GitHub repo chartgerink / 2014bump from **Distributions of p-values smaller than .05 in Psychology: What is going on?**

**2015-10-29 08:47 AM**

CHJ Hartgerink added file `submission/distributions-p_JMW mva.docx` to GitHub repo chartgerink / 2014bump to **Distributions of p-values smaller than .05 in Psychology: What is going on?**

**2015-09-21 03:46 PM**

CHJ Hartgerink added file `archive/library.bib` to GitHub repo chartgerink / 2014bump to **Distributions of p-values smaller than .05 in Psychology: What is going on?**

**2015-09-21 09:01 AM**




CHJ Hartgerink removed file `Info peek statistic.docx` from GitHub repo chartgerink / 2014bump from **Distributions of p-values smaller than .05 in Psychology: What is going on?**



# 150211FullFile\_AllStatckData\_Automatic1Tail .CSV

[Share](#)[Download](#)[View](#)[Revisions](#)

## Revisions

Version		User	Download	MD5	SHA2
ID	Date				
1	2015-04-11 06:01 PM	Michele B. Nuijten	42 	 bc59825f811af84873c9	 eb3b544a272450d354

manuscript.pdf

manuscript\_final.pdf

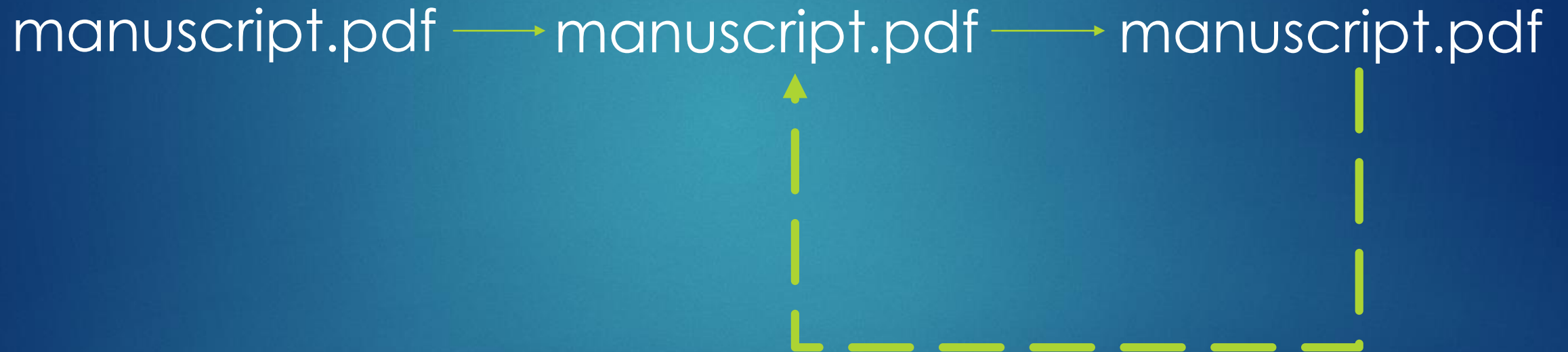
manuscript\_final2.pdf

manuscript\_final3.pdf

Timepoint 1

Timepoint 2

Timepoint 3



# How to version control

- ▶ Open Science Framework ([www.osf.io](http://www.osf.io))
  - ▶ Pro:
    - ▶ Easy to use
    - ▶ Fully free (and will stay free)
    - ▶ Collaborative, online environment
    - ▶ Optional, free, statistical consulting
    - ▶ Pre-registration possible
    - ▶ Links to services such as Dropbox (effortless version control!)
  - ▶ Cons:
    - ▶ Servers located in the United States (watch out with sensitive data! European rights are not protected in the US)



# How to version control

- ▶ Want more advanced options?
  - ▶ E.g. Github
  - ▶ Talk to me about these if you're interested!
  - ▶ More technical, but also more control

## Recent activity

All times displayed at +0100 UTC offset.

- 2016-01-05 11:16 AM** CHJ Hartgerink updated file [analysis/master.R](#) to GitHub repo [chartgerink / 2014bump](#) in [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-12-20 10:37 AM** CHJ Hartgerink updated file [analysis/master.R](#) to GitHub repo [chartgerink / 2014bump](#) in [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-12-18 05:33 PM** CHJ Hartgerink updated file [analysis/master.R](#) to GitHub repo [chartgerink / 2014bump](#) in [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-11-30 11:51 AM** CHJH updated file [analysis/master.R](#) to GitHub repo [chartgerink / 2014bump](#) in [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-11-30 11:51 AM** CHJH added file [desktop.ini](#) to GitHub repo [chartgerink / 2014bump](#) to [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-11-16 10:00 AM** C.H.J. Hartgerink updated the license of [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-10-29 09:50 AM** CHJ Hartgerink removed file [submission/distributions-p\\_JMW mva.docx](#) from GitHub repo [chartgerink / 2014bump](#) from [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-10-29 08:47 AM** CHJ Hartgerink added file [submission/distributions-p\\_JMW mva.docx](#) to GitHub repo [chartgerink / 2014bump](#) to [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-09-21 03:46 PM** CHJ Hartgerink added file [archive/library.bib](#) to GitHub repo [chartgerink / 2014bump](#) to [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)
- 2015-09-21 09:01 AM** CHJ Hartgerink removed file [Info peek statistic.docx](#) from GitHub repo [chartgerink / 2014bump](#) from [Distributions of p-values smaller than .05 in Psychology: What is going on?](#)



This repository Search

Pull requests Issues Gist



CenterForOpenScience / rpp

Unwatch 6

Star 3

Fork 3

Code

Issues 0

Pull requests 0

Wiki

Pulse

Graphs

Branch: master

Commits on Dec 30, 2015



Revert "nore. code"

FredHasselmann committed 20 days ago



215bea7



nore. code

FredHasselmann committed 20 days ago



c488c54



Commits on Nov 2, 2015



add code for vieland response reanalysis

chartgerink committed on 2 Nov 2015



bdd77e3



Commits on Sep 11, 2015



Delete rpp\_data.csv

RobbievanAert committed on 11 Sep 2015



c56abb3



Commits on Sep 10, 2015



Added columns to data

RobbievanAert committed on 10 Sep 2015



c755df6



Commits on Sep 4, 2015



Test Github auto .zip

FredHasselmann committed on 4 Sep 2015



a109387



Can you provide a  
documented account of  
your process?



# Sharing research

- ▶ Sharing with co-author(s)
  - ▶ MANDATORY for data package ( $\geq 1$ )
  - ▶ Good practice
    - ▶ Check each other's analyses
    - ▶ Check reported results for rounding errors (1/2 papers in psych contain error)
- ▶ Sharing with everyone
  - ▶ Good practice → allows everyone to verify results
  - ▶ Becoming the new standard
  - ▶ Public preservation of research
- ▶ Both possible with OSF

Make Private

Public




0



# Distributions of p-values smaller than .05 in Psychology: What is going on?

Contributors: [C.H.J. Hartgerink](#), Marcel A.L.M. van Assen, R.C.M. van Aert, Michele B. Nuijten, Jelte Wicherts

Date created: 2014-11-12 10:40 AM | Last Updated: 2016-01-05 11:16 AM

Category: Project 

Description: Here we investigate whether the bump below .05 as proposed by Masicampo and Lalande (2012) and Leggett et al. (2013) is present in 258,105 datapoints extracted from flagship journals in psychology.

License: CC0 1.0 Universal

# Distributions of p-values smaller than .05 in Psychology: What is going on?

Make Private

Public




0



Contributors: [C.H.J. Hartgerink](#), Marcel A.L.M. van Assen, R.C.M. van Aert, Michele B. Nuijten, Jelte Wicherts

Date created: 2014-11-12 10:40 AM | Last Updated: 2016-01-05 11:16 AM

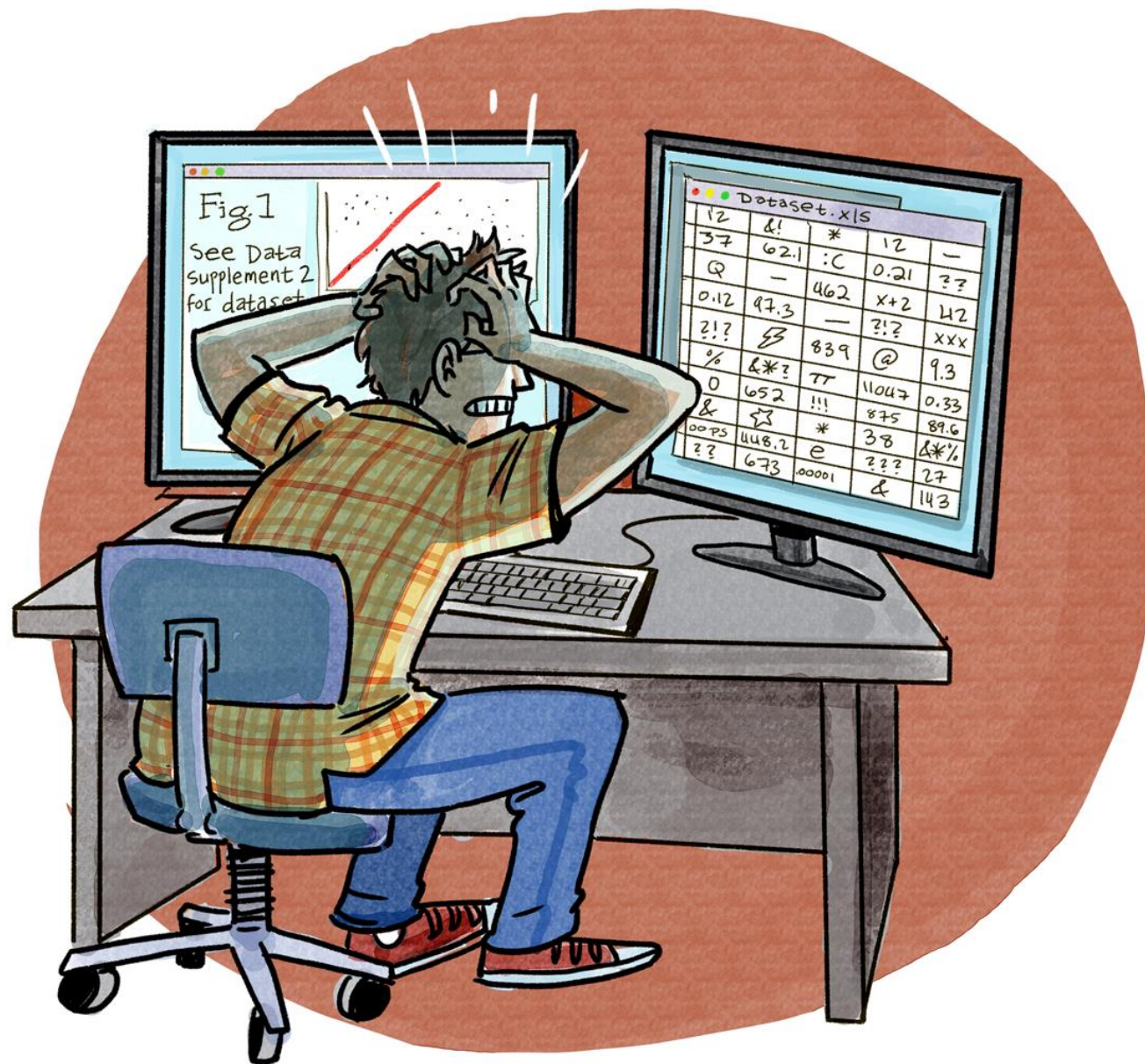
Category: Project 

Description: Here we investigate whether the bump below .05 as proposed by Masicampo and Lalande (2012) and Leggett et al. (2013) is present in 258,105 datapoints extracted from flagship journals in psychology.

License: CC0 1.0 Universal

Are the research files  
directly available to  
someone else if you were  
to fall ill?





# Reproducible

- ▶ ALL steps from raw data to results
- ▶ Use the syntax for everything
  - ▶ Reading in dataset
  - ▶ Variable transformations
  - ▶ Creating sum scores
  - ▶ How you handle missing cases
  - ▶ Etc.
- ▶ Make it as easy as possible

# Reproducible

- ▶ Leave the least amount unclear
  - ▶ Using excel to calculate some new variable? ☹
  - ▶ Using graphical interface for SPSS without pasting the syntax? ☹
  - ▶ Not mentioning what software you used? ☹
  - ▶ Not caring about these things? ☹
- ▶ Remember, you are shooting yourself (!) in the foot

# Reproducible

```
GET  
  FILE='C:\Data\Full_Dataset_De-Identified.sav'.  
DATASET NAME DataSet1 WINDOW=FRONT.
```



```
UNIANOVA anchoring1a BY expgender  
  /METHOD=SSTYPE(3)  
  /INTERCEPT=INCLUDE  
  /CRITERIA=ALPHA(0.05)  
  /DESIGN=expgender.
```



# Reproducible

```
GET  
FILE='C:\Data\Full_Dataset_De-Identified.sav'.  
DATASET NAME DataSet1 WINDOW=FRONT.
```



X

```
UNIANOVA anchoring1a BY expgender  
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=expgender.
```

# Reproducible

GET

FILE='C:\Data\Full\_Dataset\_De-Identified.sav'.  
DATASET NAME DataSet1 WINDOW=FRONT.

COMPUTE anchoring\_sum = anchoring1a + anchoring1b.  
EXECUTE.

UNIANOVA anchoring BY expgender  
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=expgender.



# Reproducible

GET

FILE='C:\Data\Full\_Dataset\_De-Identified.sav'.  
DATASET NAME DataSet1 WINDOW=FRONT.

~~COMPUTE anchoring\_sum = anchoring1a + anchoring1b.~~  
~~EXECUTE.~~

UNIANOVA anchoring BY expgender  
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=expgender.



# Reproducible

GET

```
FILE='C:\Data\Full_Dataset_De-Identified.sav'.  
DATASET NAME DataSet1 WINDOW=FRONT.
```

```
COMPUTE anchoring_sum = anchoring + anchoring1b.  
EXECUTE.
```

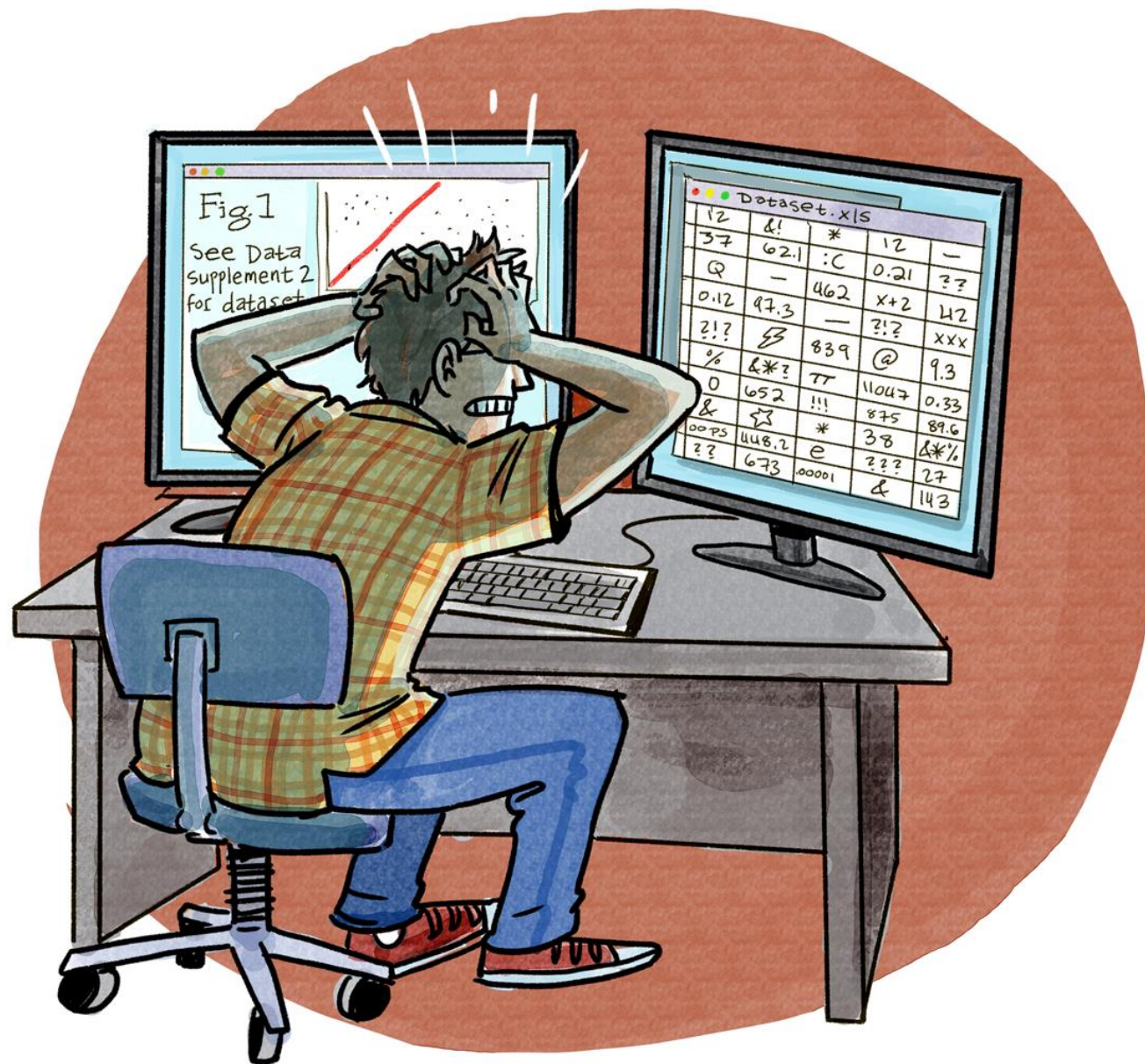
**ERROR**

X

UNIANOVA anchoring BY expgender

```
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=expgender.
```





# Reproducible

- ▶ Future is dynamic documents
  - ▶ E.g., knitr, Rmarkdown
- ▶ The effect of condition on the dependent variable was significant,  $F(\texttt{`r result$df1`}, \texttt{`r results$df2`}) = \texttt{`r result$fvalue`}$ ,  $p = \texttt{`r result$pval`}$ .
- ▶ The effect of condition on the dependent variable was significant  $F(1, 65) = 10$ ,  $p = .002$ .
- ▶ Interested in working with this, let me know!

# Future proof

- ▶ File formats matter
- ▶ Reading word file from 1992?

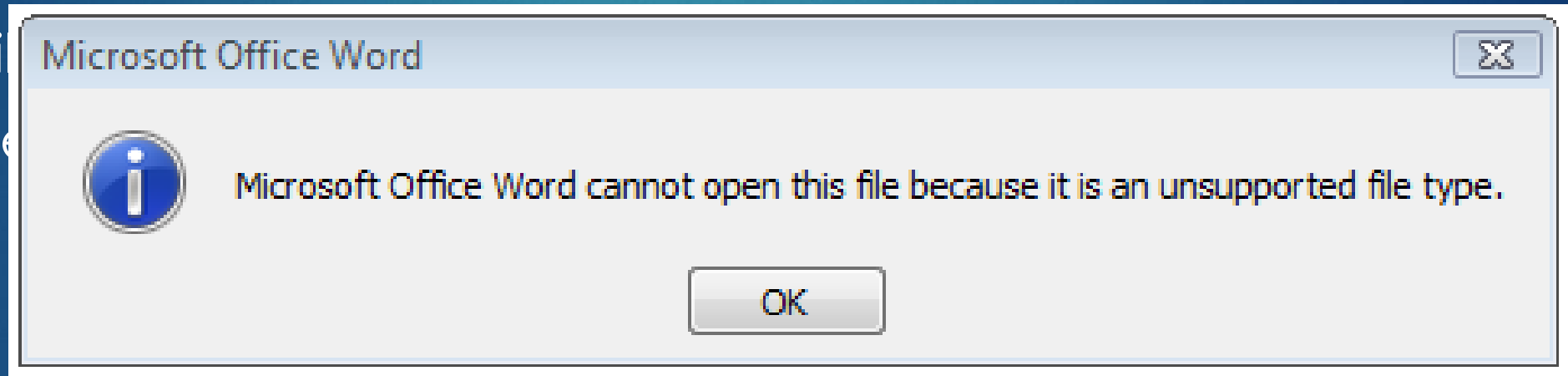
# Future proof



File



Re

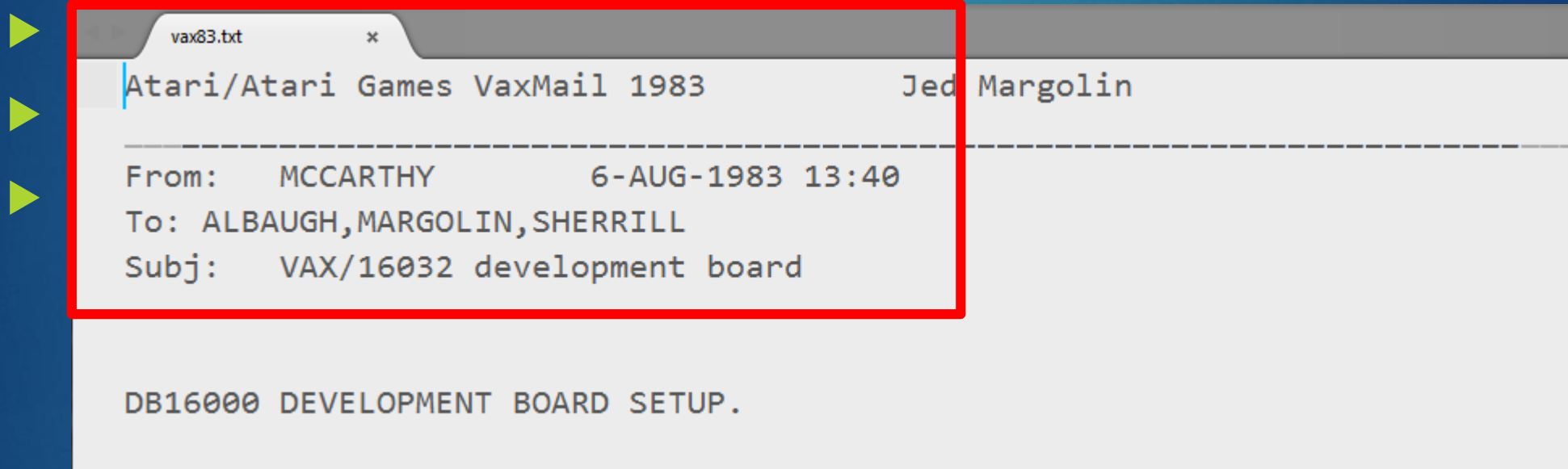




# Future proof

- ▶ File formats matter
- ▶ Reading word file from 1992?
- ▶ Reading text file from 1983? (oldest I could find)

# Future proof



# Future proof

- ▶ Proprietary file formats (.docx .xlsx .sav) not future proof
- ▶ Changes in the software make these files difficult to read!
- ▶ Companies do not make public how the format changed (“company secret” and “intellectual property”)
- ▶ Open formats always readable
  - ▶ .txt
  - ▶ .csv
  - ▶ .dat
  - ▶ .odt
- ▶ Save files in open formats

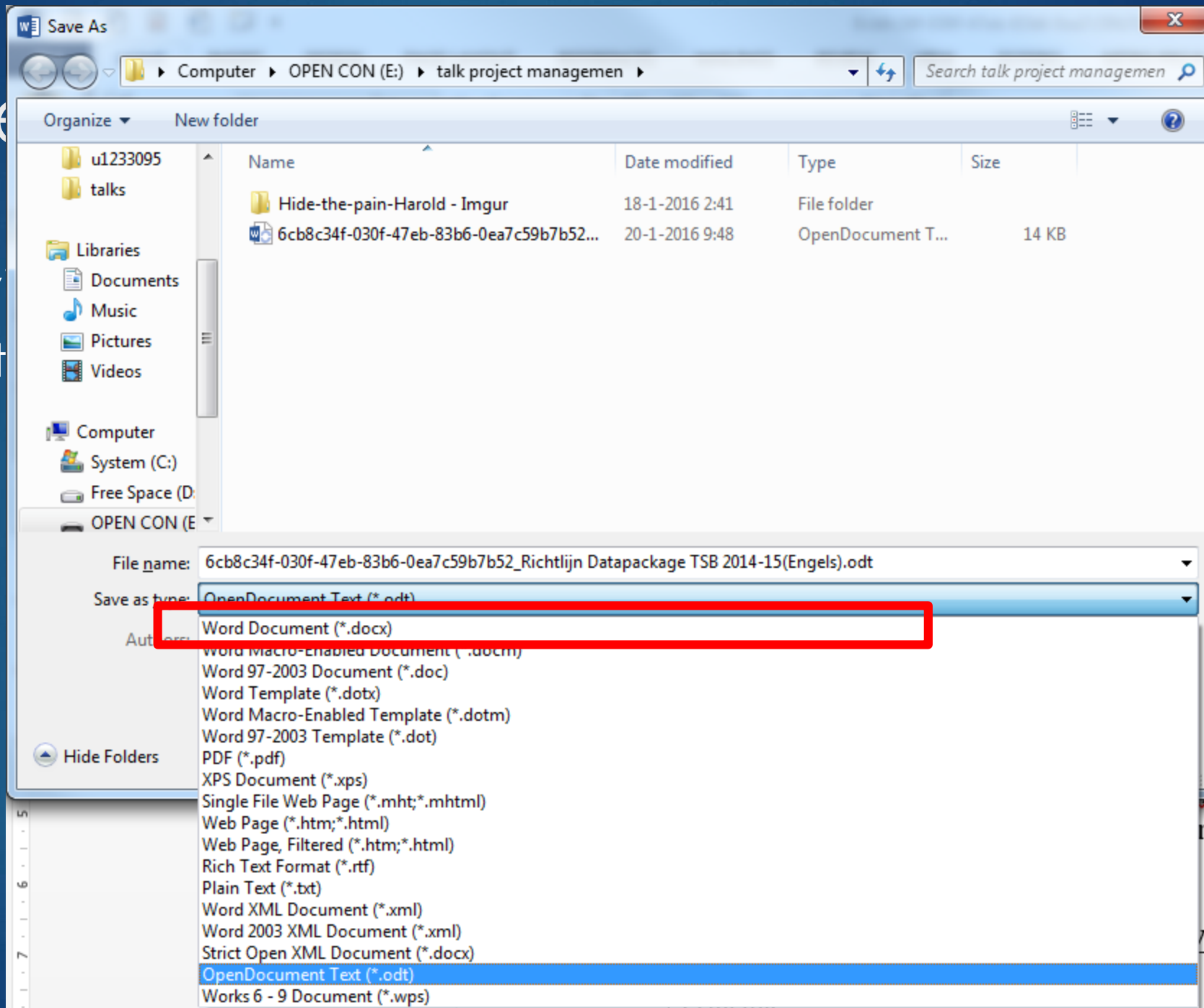
# Future proof

- ▶ Saving files in open formats requires almost no change
- ▶ Just change the file format when saving

# Future

► Save

► Just

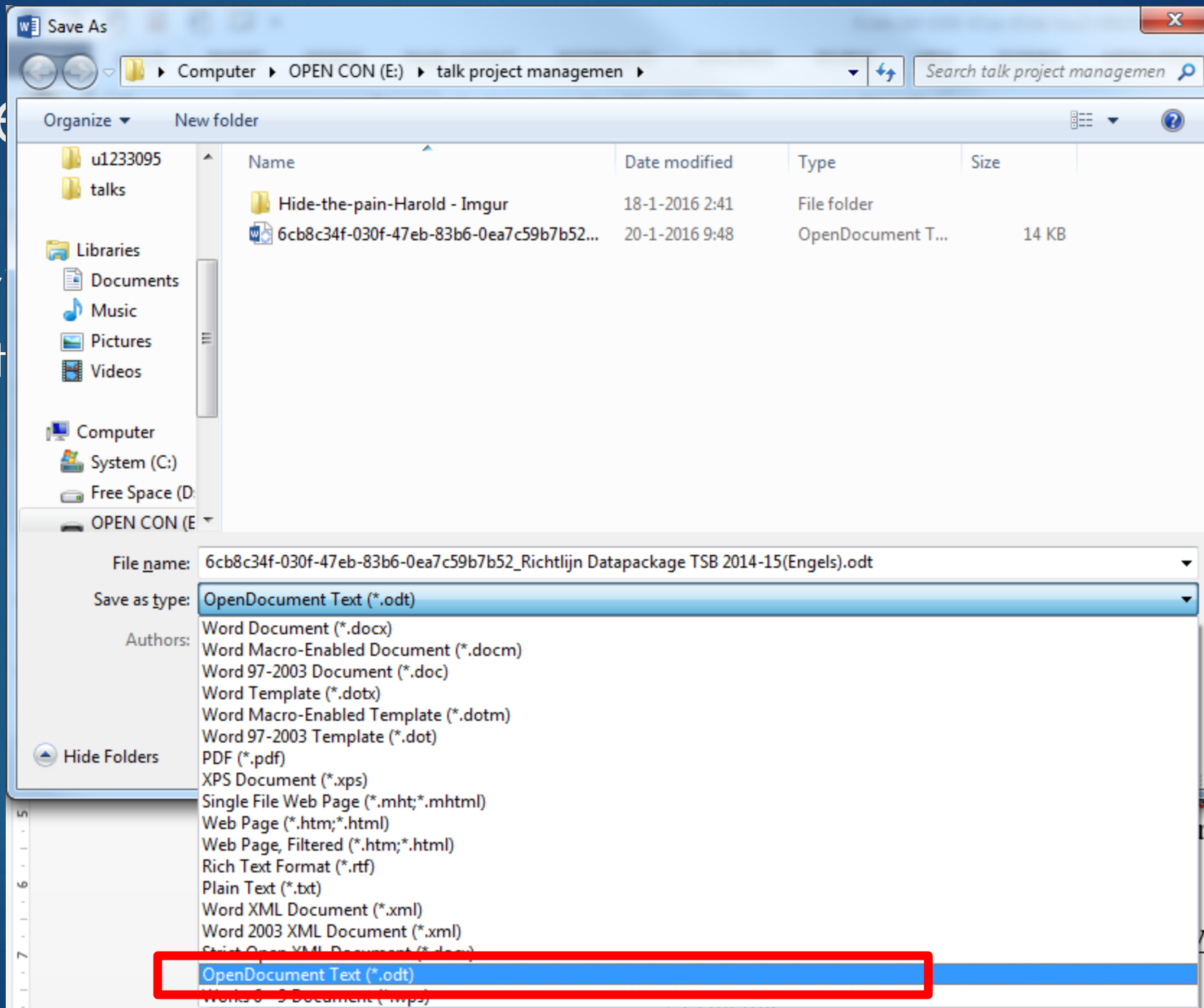




# Future

► Save

► Just



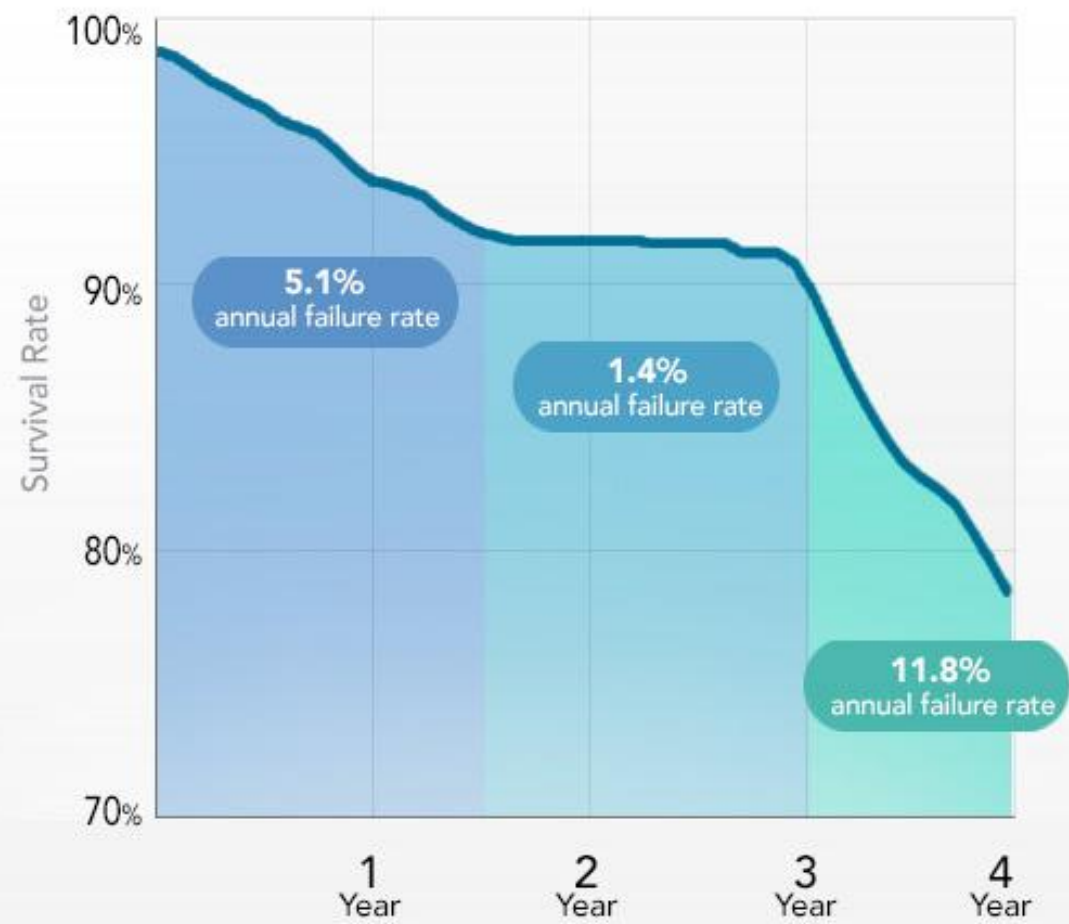
# Future proof

- ▶ Saving files in open formats requires minor change
- ▶ Just change the file format when saving
- ▶ Can still track changes, work with co-authors, etc.
- ▶ PLUS future proofed.

Can you still use these files  
in 10 years?

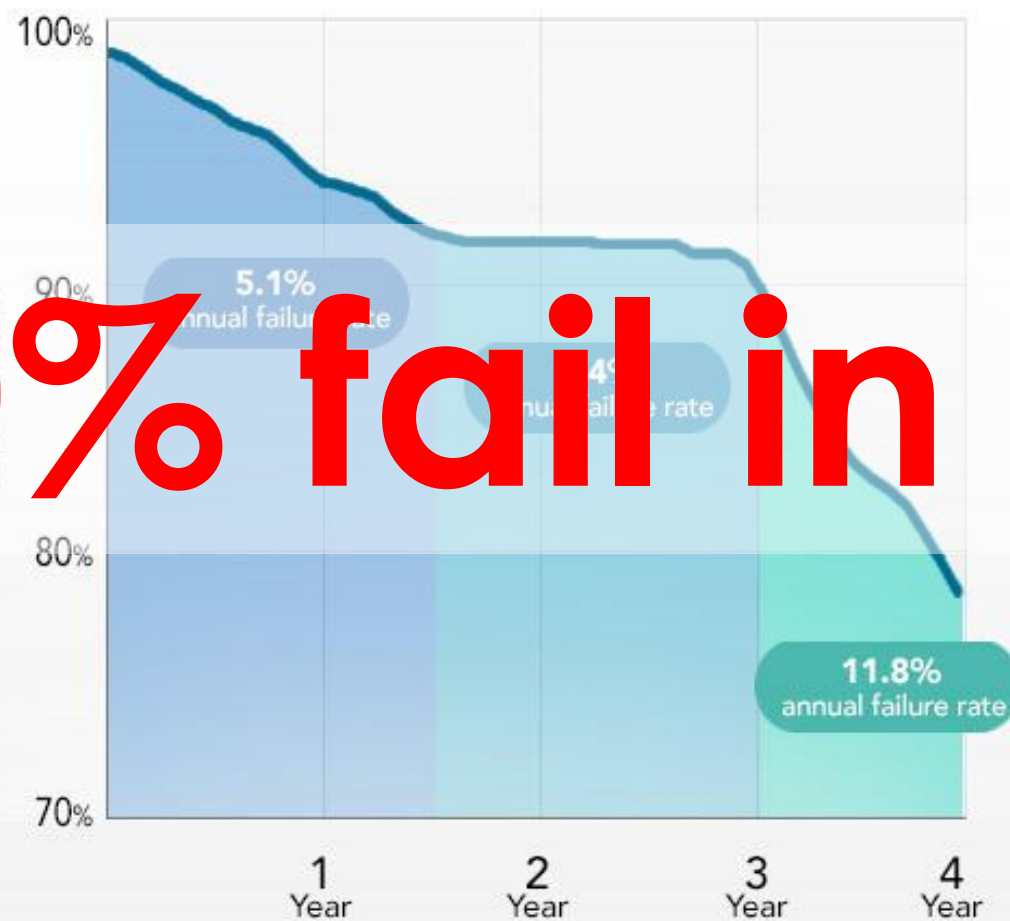
# Back-ups

- ▶ Essential for preservation
- ▶ Online + offline possibilities
- ▶ Open Science Framework helps online back up
  - ▶ Continuous: link OSF with Dropbox, Google Drive, etc.
  - ▶ Manual: save files on the OSF
- ▶ Offline backups remain important (sensitive data)





20% fail in 4y



If your hard drive is stolen,  
crashes, are you in trouble?

# Ask yourself...

- ▶ Are all research files collected in one place?
- ▶ Are the research files organized?
- ▶ Is it clear who did what in the project?
- ▶ Are results and figures directly reproducible by others?
- ▶ Are the research files available to co-author(s) or beyond?
- ▶ Are your research files backed-up?

# Data package

- ▶ Mandatory in TSB!
- ▶ Yearly audits of random sample of published papers
- ▶ If you get audited, it *\*must\** be in order
  - ▶ Option 1: Be like Harold
  - ▶ Option 2: Don't be like Harold

# Data package

## ► Eight requirements

1. Metadata
2. Raw data
3. Data storage
4. Material
5. Statistical processing
6. Processed database
7. Access
8. Retention period



# Data package

## ► Eight requirements

1. Metadata ✓
2. Raw data ✓
3. Data storage ✓
4. Material ✓
5. Statistical processing ✓
6. Processed database ✓
7. Access ✓
8. Retention period ✓



VICTORY ACHIEVED

# Resources

- ▶ Template project structure on the Open Science Framework:  
[www.osf.io/4sdn3](http://www.osf.io/4sdn3)
- ▶ Data audit guidelines  
<https://www.tilburguniversity.edu/research/social-and-behavioral-sciences/download-guideline-datapackage-tsb/>
- ▶ I am always prepared to help out with reproducibility of research
  - ▶ [chjh@protonmail.com](mailto:chjh@protonmail.com)
  - ▶ @chartgerink
  - ▶ P1.174B
  - ▶ Coffee machine