

# Research Data Management Why and How?



**Yasemin Turkeyilmaz-van der Velden, Nicolas Dintzer & Esther Plomp**

Data Stewards @ TU Delft Faculties of 3mE, TPM and TNW  
y.turkyilmaz-vandervelden@tudelft.nl, N.J.R.Dintzner@tudelft.nl,  
E.Plomp@tudelft.nl

Slides are available: <https://doi.org/10.5281/zenodo.3537598>

# Who are we?

## Data Steward @ TU Delft

[www.tudelft.nl/library/datastewardship/](http://www.tudelft.nl/library/datastewardship/)

Secure data storage, data sharing, citation



Advice



Archiving

For data management in grant proposals



Costs

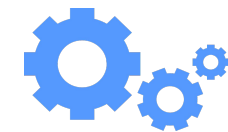


Compliance Management Plans

Advice and templates



Data



Tools

Workshops, information sessions



Training

4TU.Centre for Research Data or disciplinary repositories

With funders' and journals' policies

For data and software management

# Data

**descriptive/numerical** information considered for reference or analysis

- notebooks
- survey responses
- software and code
- measurements
  - laboratory/field equipment
- images (photographs, scans)
- audio/video recordings
- physical samples

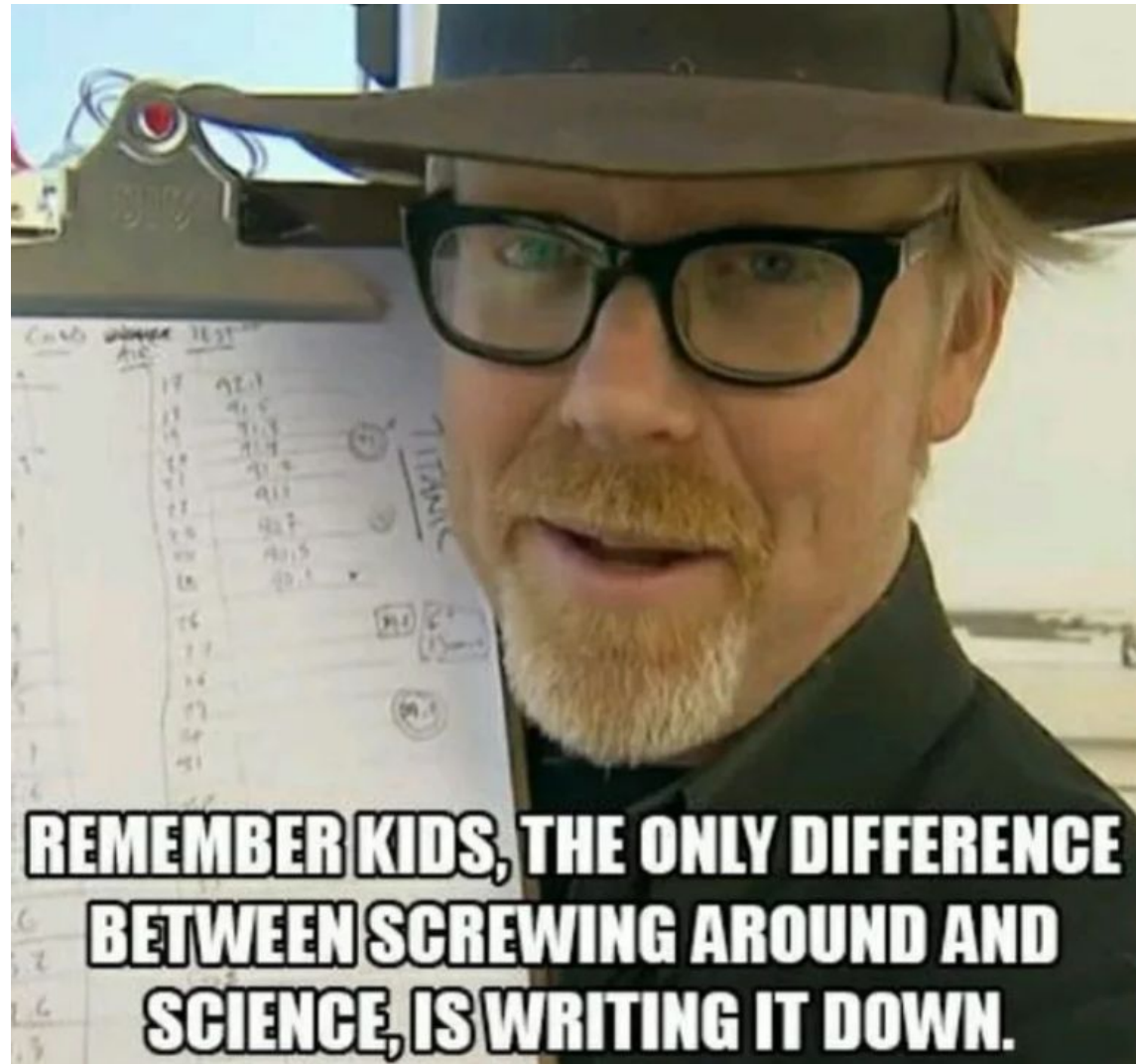
# Research Data Management

The **organisation** of data throughout the research project

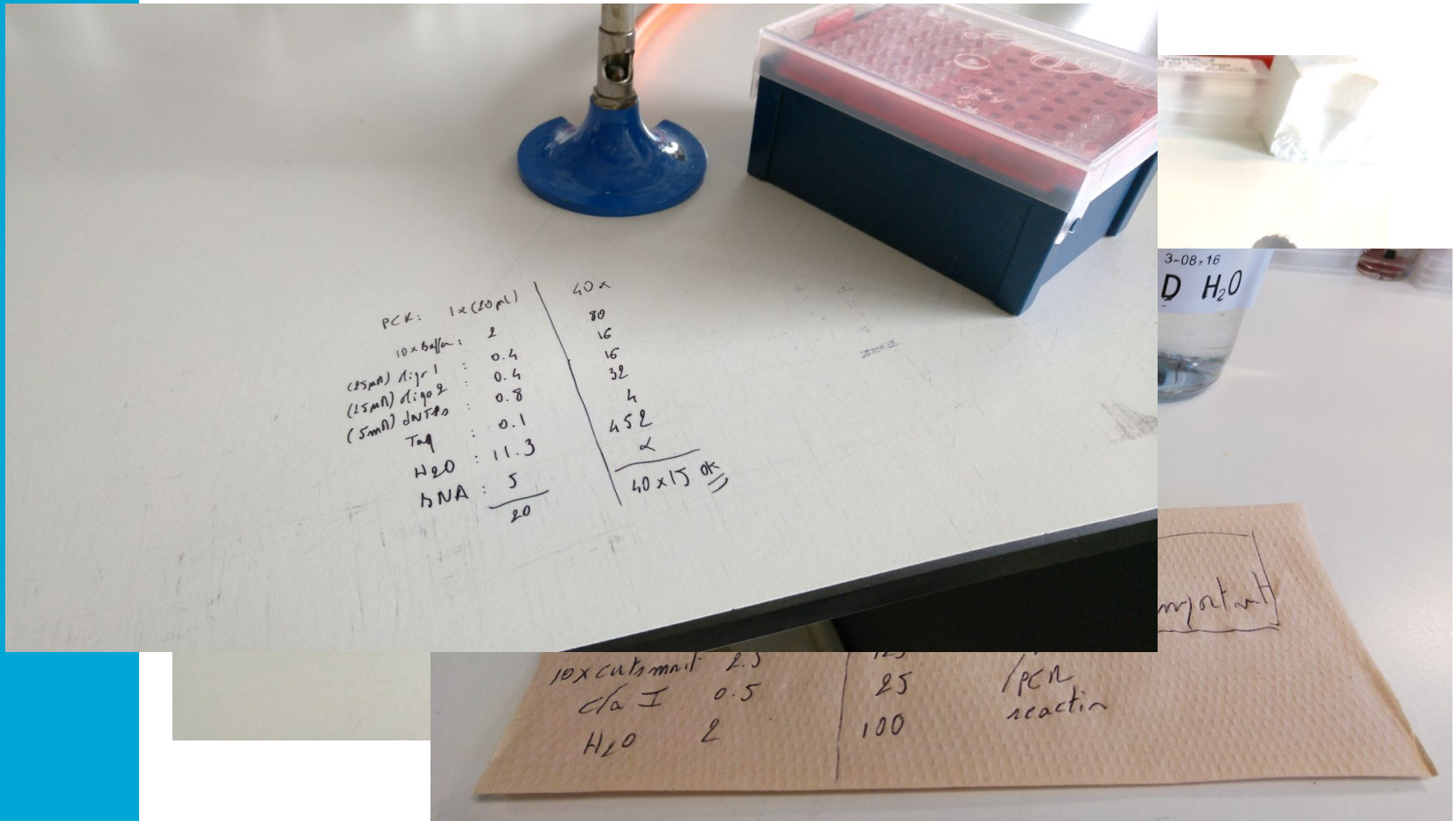
- everyday management of research data
  - storage, file naming, documentation
- preservation and sharing of data after the project is completed



For today

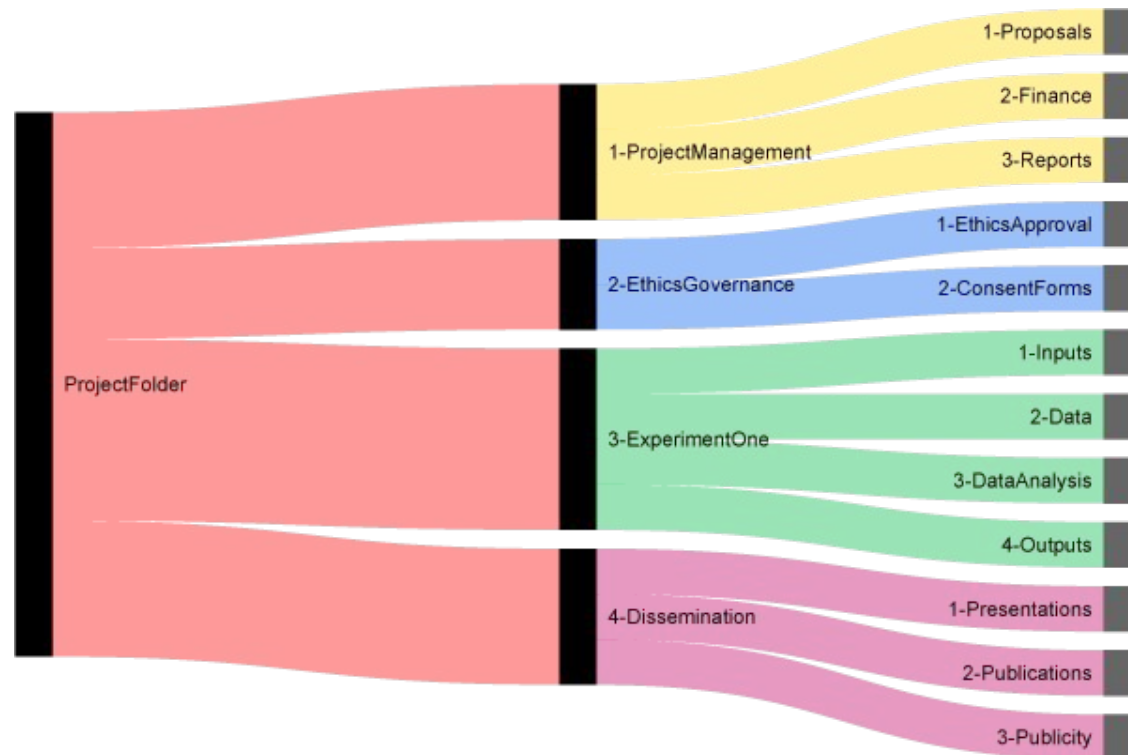


# First things first...

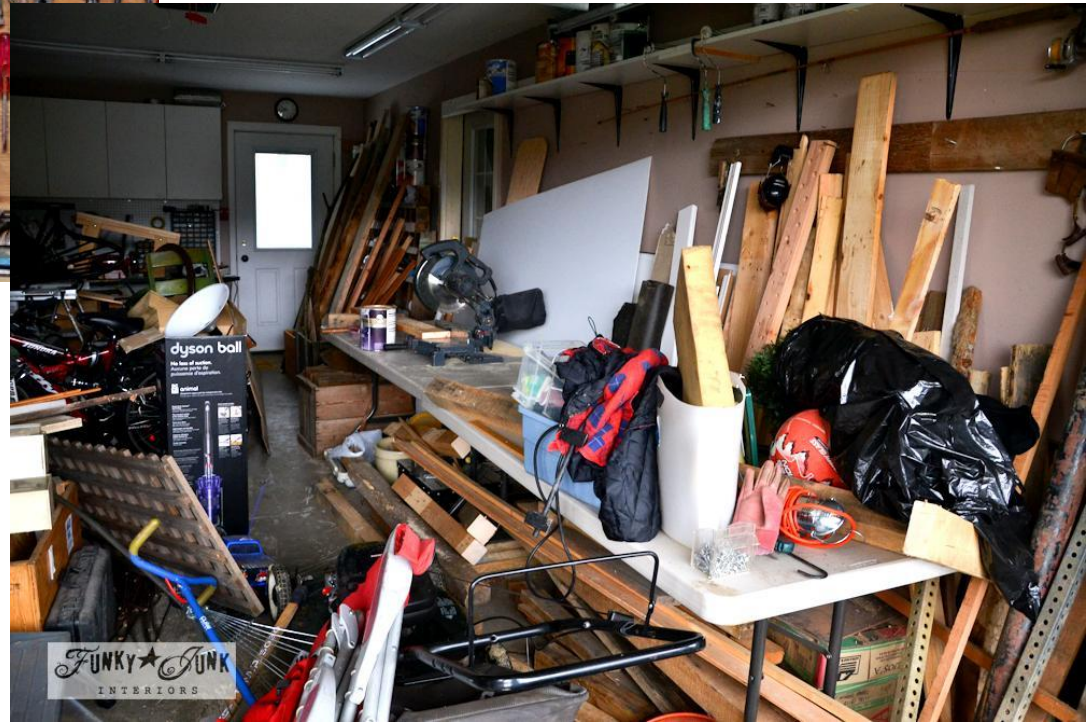


Data collection protocols – where does your data come from, and how are you keeping track of it ?

# Advice 1: don't lose your stuff, “organize it”



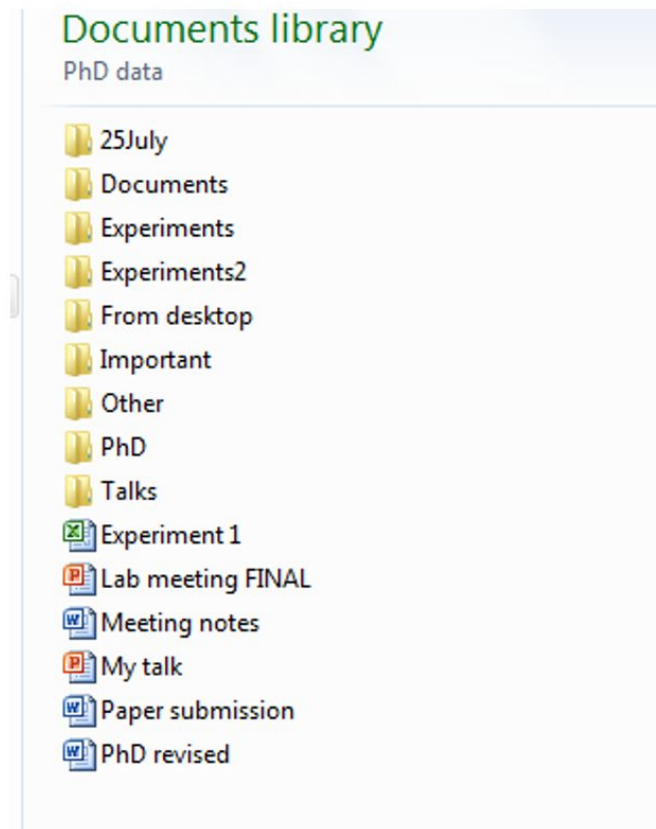
# Organized / Dumped ?



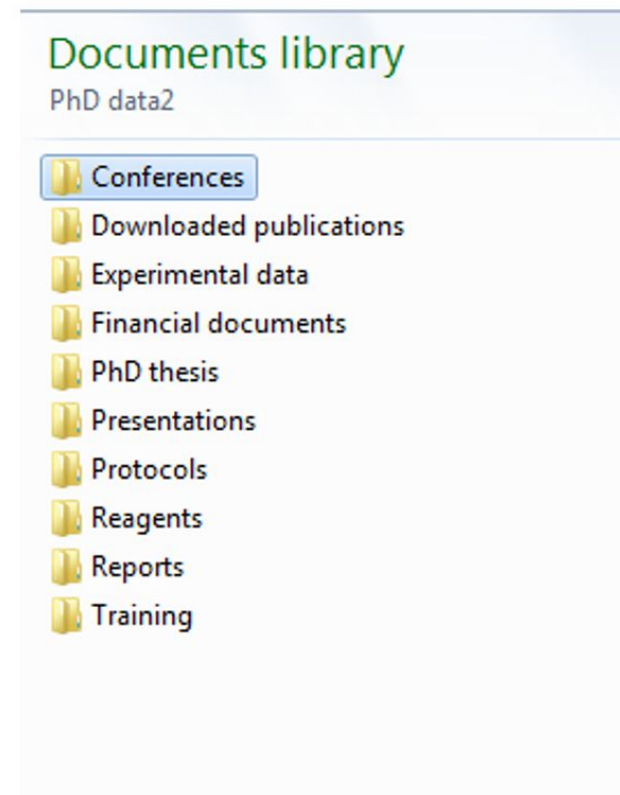


# Practical example

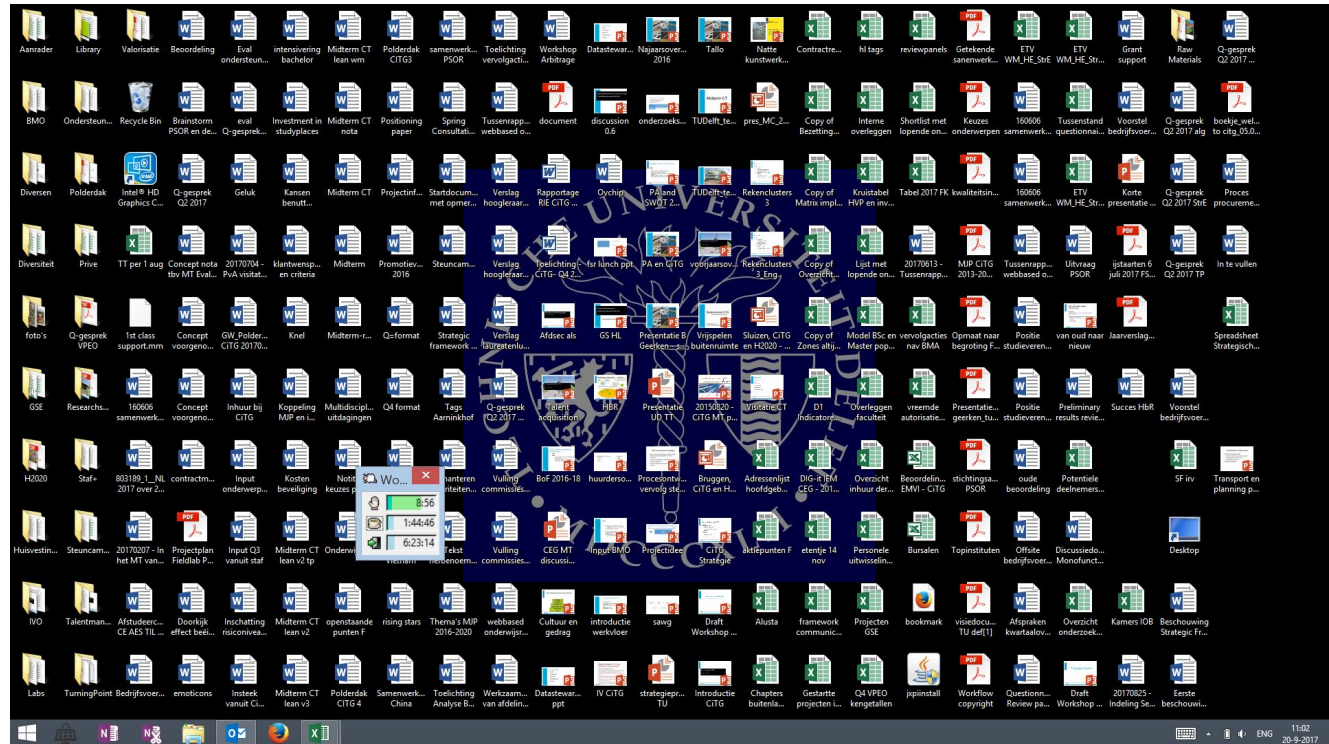
## Example A



## Example B



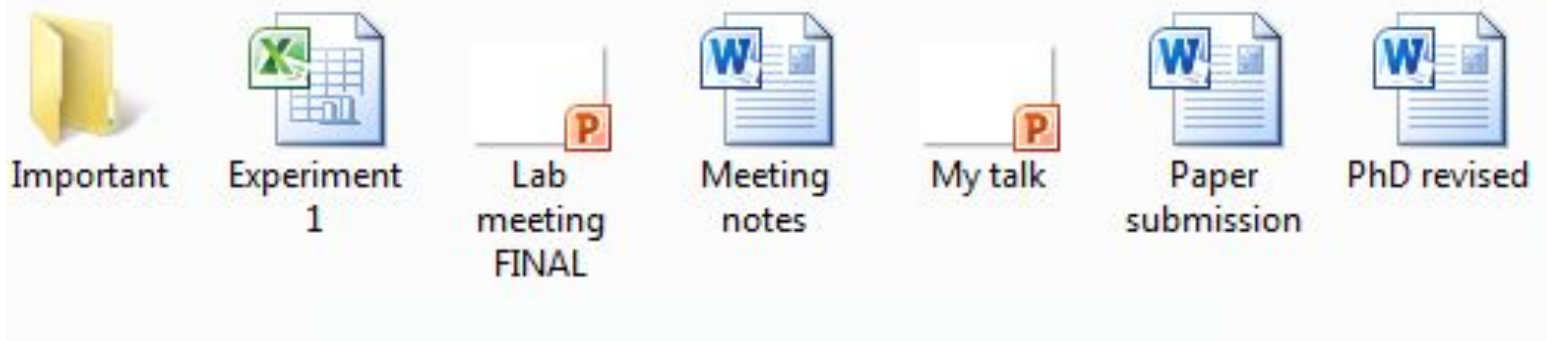
# Organized / Dumped ?



- Consistent
- Meaningful to you and your colleagues
- Allow you to find files easily
- [Project] / [Experiment] / [Instrument or Type of file] / [Date]

# File naming

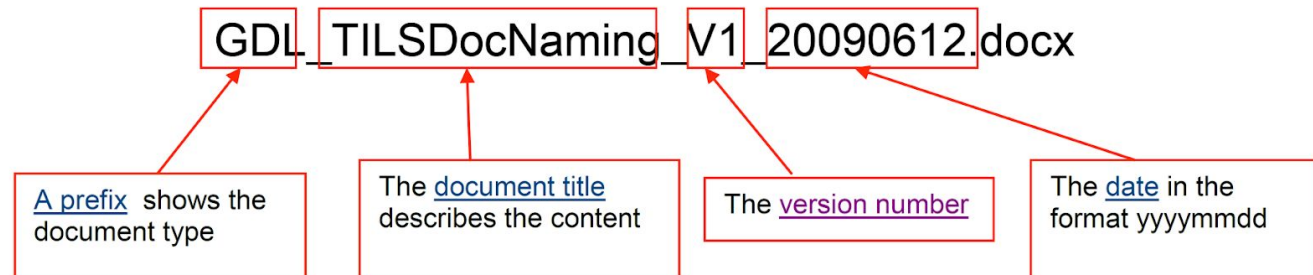
In 3 years time would you know what these are?



# Advice 2: don't lose your stuff, Name it

## TILS Document Naming Convention

Document naming for the TILS Division should follow this convention:



- Date or date range of experiment: YYYYMMDD
- File type ( “lab\_report”, “Lecture\_X-assignment-1”...)
- Researcher name/initials
- Version number of file
- Don't make file names too long
- Avoid special characters and spaces
- Include a README.txt file to explain the naming convention

# But then...

## Shedding Light on the Dark Data in the Long Tail of Science

P. BRYAN HEIDORN



Philosophy & Technology  
pp 1-23 | [Cite as](#)

### Dark Data as the New Challenge for Big Data Science and the Introduction of the Scientific Data Officer

Authors Authors and affiliations

Björn Schembera, Juan M. Durán 

[Open Access](#) | Research Article  
First Online: 13 March 2019

7 Shares 2.6k Downloads

“Dark data” is data that we know is here, but lack of knowledge about it makes it useless

# Devil is in the details

Name	Date	Location	Position
...	...	...	...

**Name** : interviewee

**Date**: date of the interview

**Location**: company where the interview took place

**Position**: position of the interviewee in the company

Name: bob

Date: 10/10/2010

Location: ABN Paris

Position: cyber security manager

**Name** : car owner

**Date**: date of the recorded accident

**Location**: location of the accident

**Position**: position on the road of the car at the moment of the accident

Name: bob

Date: 10/10/2010

Location: A2 - km 55

Position: middle lane

# Vocabulary: how to say 'female'

18-day pregnant females	Female (lactating)	Individual female	Worker caste 'female'
2 yr old female	Female (pregnant)	Igb*cc females	Sex female
400 yr. Old female	Female (outbred)	Mare	Female, other
Adult female	Female parent	Female (worker)	Female child
Asexual female	Female plant	Monosex female	<b>Femal</b>
Castrate female	Female with eggs	Ovigerous female	3 female
Cf.female	Female worker	Oviparous sexual females	Female (phenotype)
Cystocarpic female	Female, 6-8 weeks old	Worker bee	Female mice
Dikaryon	Female, virgin	Female enriched	Female, spayed
Dioecious female	Female, worker	Pseudohermaphroditic	<b>Femlale</b>
Diploid female	Female(gynocious)	<b>female</b>	Metafemale
F	<b>Femele</b>	<b>Remale</b>	Sterile female
Famale	Female, pooled	Semi-engorged female	Normal female
<b>Femail</b>	Femalen	Sexual oviparous female	Sf
Female	Females	Sterile female worker	Vitellogic replate female
Female – worker	Females only	Strictly female	Worker
Female (alate sexual)	Gynocious	Tetraploid female	Hexaploid female
Female (calf)	Healthy female	Thelytoky	Female (f-o)
Hen	Probably female (morphology)	Female (gynocious)	

Slide adapted  
from Christine  
Staiger  
<https://doi.org/10.5281/zenodo.2585691>

# Advice 3: don't lose you stuff, document it!

- Meta data – data about your data
- Spreadsheet example – for each column
  - What is it supposed to contain ?
  - Format / units ?
  - Vocabulary ?
- Located in:
  - The spreadsheet itself
  - README file
- For spreadsheets, scripts / code , ...

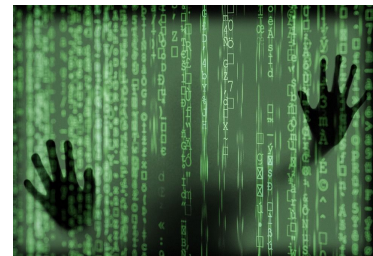




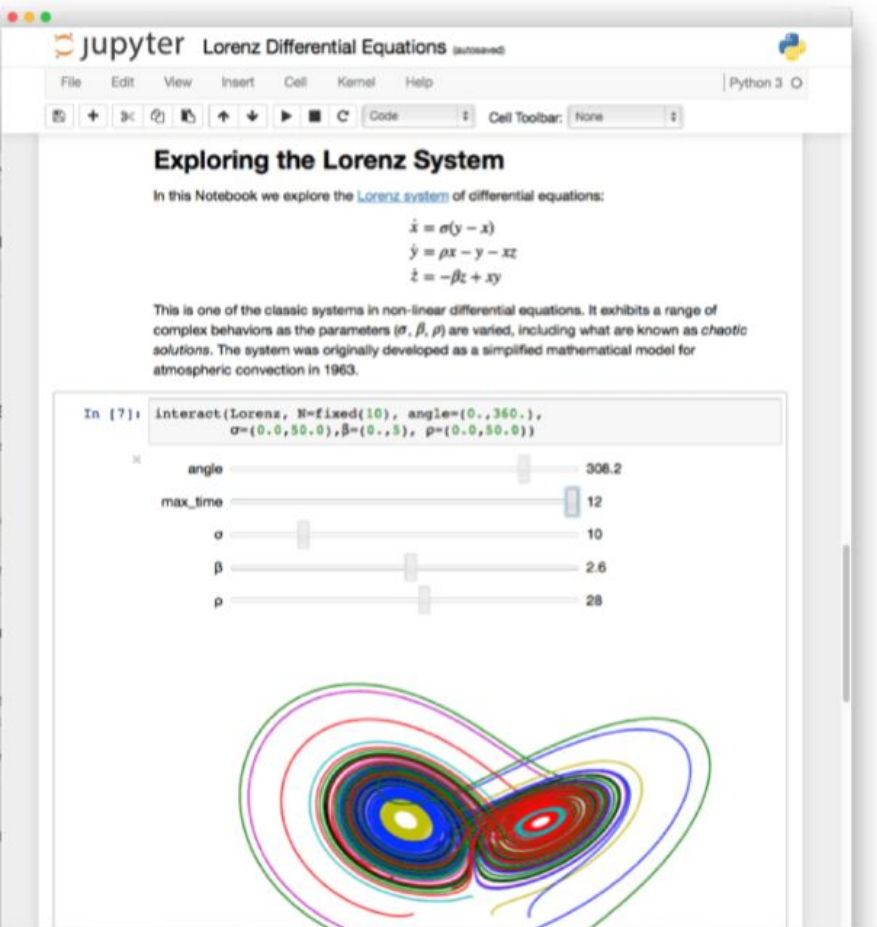
# The special case of software code

<https://github.com/fogleman/Minecraft/blob/master/main.py#L195>

- File name
  - Should help you to know what is inside !
- Function:
  - Name : what it does
  - Document the function (more info on what it does)
  - Document its input and output
- Variable names:
  - **T is time or temperature ?**
  - What it is - be clear!
  - x,y,z ?



# Open notebooks such as Jupyter & Rstudio



Open Notebooks are documents that contain equations, visualisations, narrative text and live code that can be executed independently and interactively, with output visible immediately beneath the input.

They bring together analysis descriptions and results, which can be executed to perform the data analysis in real time.

Added value:

- Transparency in the analysis of the data
- Reproducibility
- Documentation of the entire workflow

# So far, so good ?

- Don't lose your stuff!
  - Organize it
  - Name it
  - Document it
- ... Are we good ?

# CASH REWARD

for returning my lost backpack



www.adventure.com

- Black [AK] Burton Rucksack
- Lost on Friday 15. July at 8 pm in the Panton Arms pub 43, Panton St. Cambridge
- Containing a laptop (white MacBook), a black external hard drive and scientific research documents

The external hard drive is **VERY** important to me as it contains 5 years of research data which are crucial for my PhD thesis!!!

If you found it, I would be extremely grateful if you could return it to the Panton Arms or contact me on: [redacted]@cam.ac.uk

Thank you!!

## Large fire at the Faculty of Architecture Delft

editorial staff

10.00 a.m.: A fierce fire broke out this morning at the faculty of architecture at TU Delft. From our office we can see big black pillars of smoke. There are no casualties reported.



# Would you lose data if...?

- your laptop/notebook got stolen or lost
- your lab burnt down
- you lost your USB stick
- your portable hard drive got damaged
- your files on Dropbox / Googledrive disappeared

## FOR MY LOST LAPTOP

I am a Rutgers Chemistry 5<sup>th</sup> year PhD student. On April 19<sup>th</sup> afternoon, my LENOVO THINKPAD T420S laptop was stolen from room 203 of Wright-Rieman building. If you stole my laptop and now you are reading this letter, I would like to say that you can keep the computer and I would like to pay you money for my data under D drive. The data is my FIVE-YEAR work. I really need the data under the D drive, there is a folder named RESEARCH, under RESEARCH folder, there is a THESIS folder. I only need that folder for my thesis defense, which is coming very soon. I would like to pay you \$1000 and use whatever way you offer to send you the money. The price is



original slide  
by Marta  
Teperek

# You will survive

## THE FOUR STAGES OF DATA LOSS

DEALING WITH ACCIDENTAL DELETION OF MONTHS OF HARD-EARNED DATA



[www.phdcomics.com](http://www.phdcomics.com)

But barely...

# To avoid data loss:

- Backup your data **regularly and preferably automatically**
  - Create, at a **minimum, 2 copies of your data**
  - Store data at **multiple trusted locations**
  - Use **reliable backup solutions**
- 
- Avoid data storage on hard disks, USB's, and personal computers without backup

# Not all storage born are equal

“Sensitive data”



Personal data, company data, ...



# Always read the small print...

## Google services Terms of Use:

When you upload, submit, store, send or receive content to or through our Services, you give Google (and those we work with) a worldwide license to use, host, store, reproduce, modify, create derivative works (such as those resulting from translations, adaptations or other changes we make so that your content works better with our Services), communicate, publish, publicly perform, publicly display and distribute such content. The rights you grant in this license are for the limited purpose of operating, promoting, and improving our Services, and to develop new ones. This license continues even if you stop using our Services (for example, for a business listing you have added to

# Experimental notes

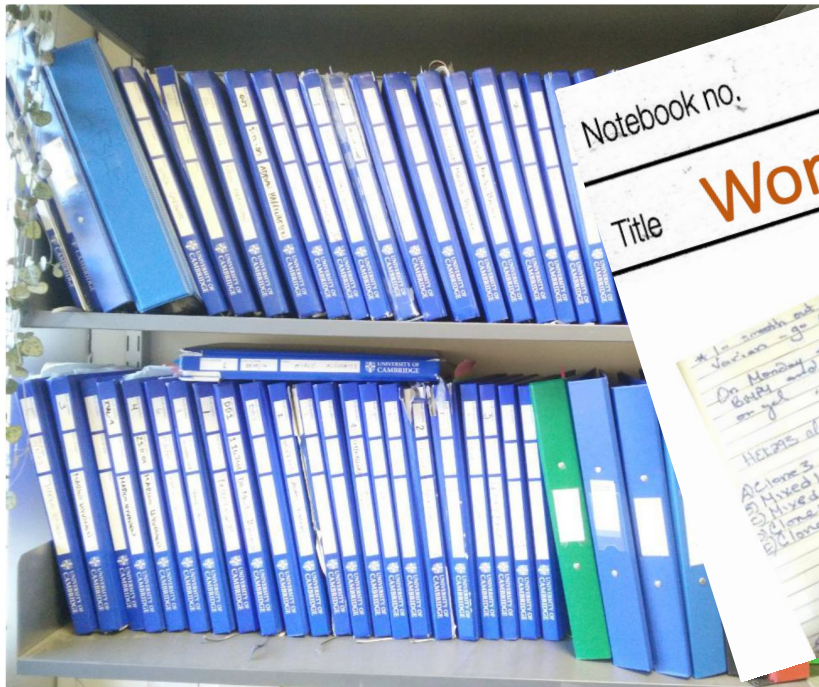
Notebook no.	Date 13 January 2017	
Title <b>The reality of today</b>		Continue

Notebook no.

Date 13 January 2017

Title

**Work for Sherlock Holmes**



At 10:30 smooth out peak line go to lamp installation  
 Variation - go to 'set up'  
 On Monday -> fully used of each division of  
 battery auto speed use to solve the problem that  
 on gel

AC clones	10:1 (25)	room
Mixed 1	10:1 (25)	room
Mixed 2	25:0:25	room
Mixed 3	25:0:25	room
AC clones 2	25:0:25	room

- none appear to have any activity  
 - Combine testing classes  
 Start all BHP4 students in Friday until Monday

MW 17 - 8:16  
 Fructose 18-21 were added  
 " 23-26 were added  
 This gel was carried out after  
 old the addition of 50 mM PGE  
 L.A. reagent pH 8.0. The dialysis was carried out  
 dialysis under 2 changes 0.5 liter vs 5 liter  
 exchange occur in 6 PM area. 2.5 liter vs 5 liter  
 The pellet affected when doctor is colour  
 gel at 25°C then when exposed at  
 then is thought to compare to the lanes

Presentation by Dr Marko Hyvonen

<https://doi.org/10.17863/CAM.7217>

# Electronic Lab Notebooks



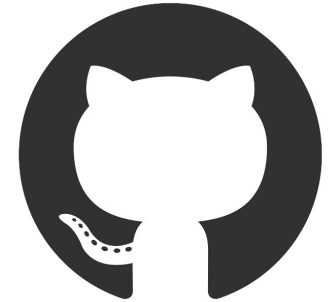
- Digital documentation, categorization and linking of
  - Raw, intermediate and final data
  - Experimental and measurement parameters
  - Physical samples
- Searchable
- Traceable (version control) & fraud-proof

Questions about the ELN trial? Contact Esther or Yasemin

# Storage space your work

- Course work: your choice - be safe
- When in doubt:
  - Ask to your supervisor
  - Go to your data steward

# Version control software



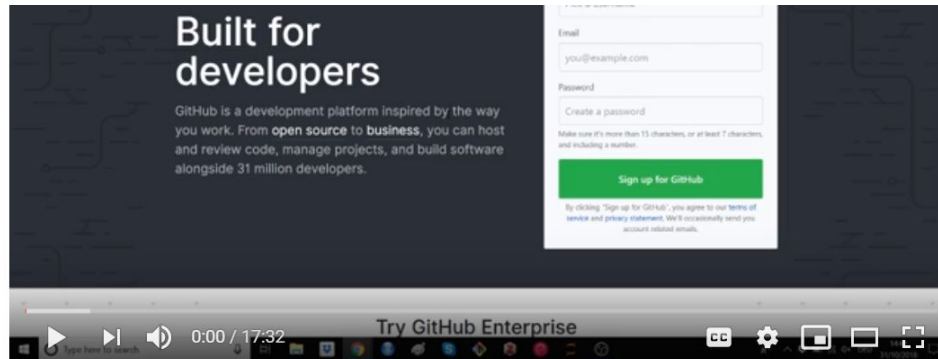
Commits on May 6, 2019

Update _config.yml EstherPlomp committed 9 days ago ✓	Verified		5d7c79c	
Update index.md EstherPlomp committed 9 days ago ✓	Verified		8d9f3c6	

Commits on Apr 17, 2019

Update index.md EstherPlomp committed 28 days ago ✓	Verified		0b86713	
--------------------------------------------------------	----------	--	---------	--

# Working with GitHub



Module 5, Task 1: How to set up a repository on GitHub

500 views

👍 7    🗨️ 0    ➦ SHARE    ≡+ SAVE    ...

<https://www.youtube.com/watch?v=AnftV9HBPSc>



<https://software-carpentry.org/lessons/>

## Lesson

The Unix Shell

Version Control with Git

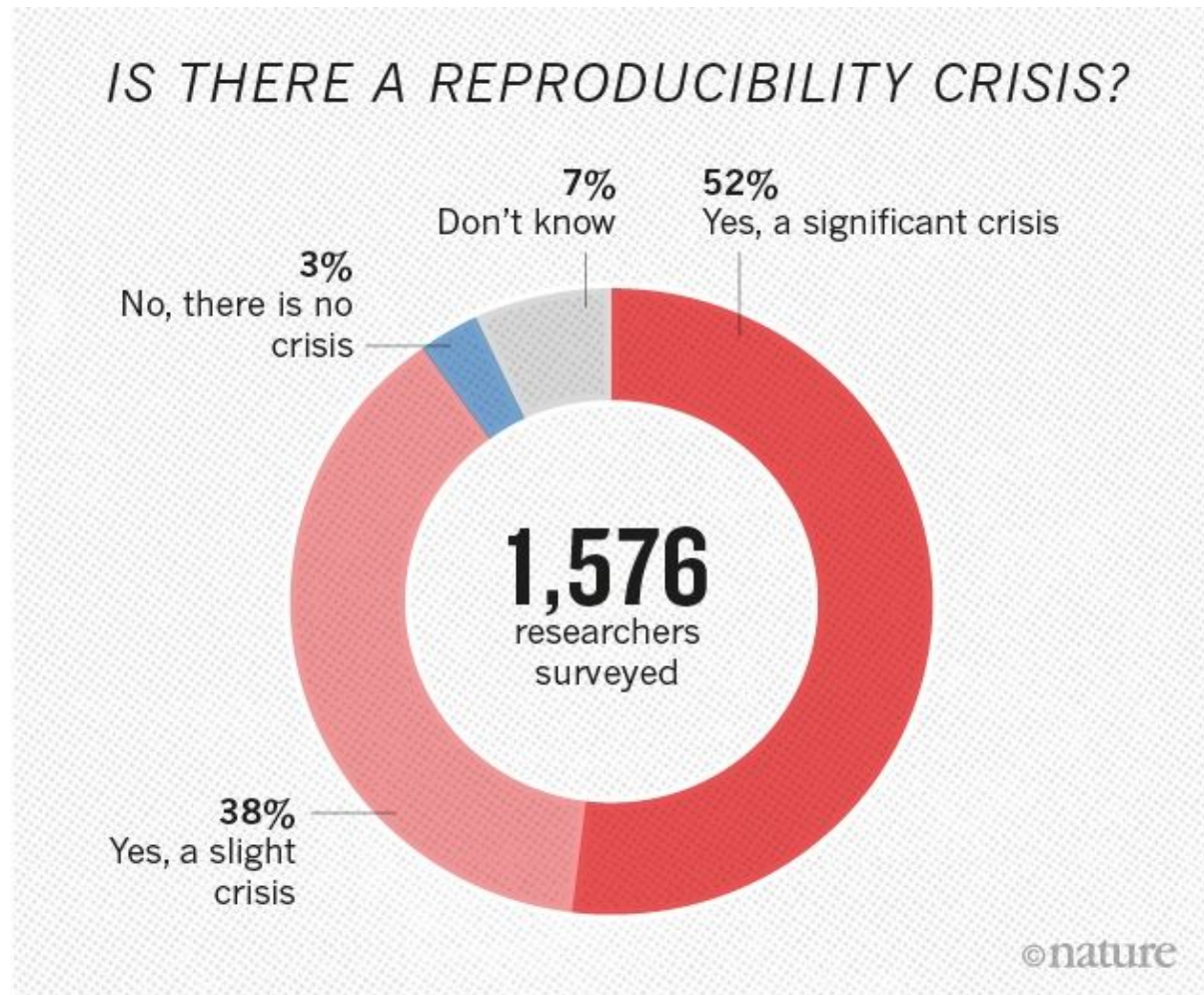
Programming with Python

Plotting and Programming in Python

Programming with R

R for Reproducible Scientific Analysis

# More challenges ahead

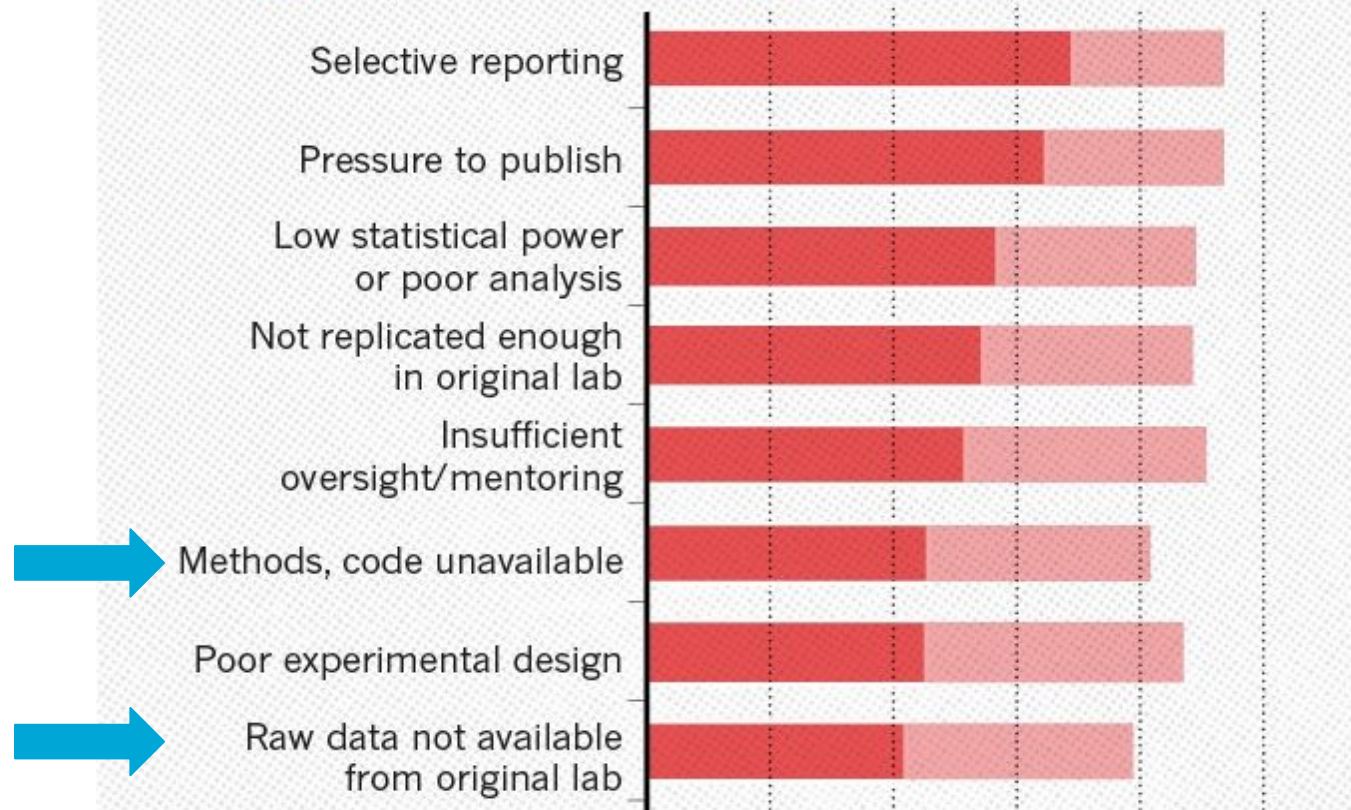


Nature 533, 452–454 (26 May 2016) doi:10.1038/533452a

# WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.

● Always/often contribute    ● Sometimes contribute



Nature 533, 452–454 (26 May 2016) doi:10.1038/533452a



# Datasets available 'on request' are not available

Current Biology 24, 94–97, January 6, 2014 ©2014 Elsevier Ltd All rights reserved <http://dx.doi.org/10.1016/j.cub.2013.11.014>

**Report**

## **The Availability of Research Data Declines Rapidly with Article Age**

- Data availability decreases by **17% per year**
- Chance of email address working decreases by **7% per year**

<http://dx.doi.org/10.1016/j.cub.2013.11.014>

# Datasets available 'on request' are not available

Current Biology 24, 94–97, January 6, 2014 ©2014 Elsevier Ltd All rights reserved <http://dx.doi.org/10.1016/j.cub.2013.11.014>

**Report**

## **The Availability of Research Data Declines Rapidly with Article Age**

- Data availability decreases by **17% per year**
- Chance of email address working decreases by **7% per year**

What's the alternative to sharing 'on request'?

<http://dx.doi.org/10.1016/j.cub.2013.11.014>

# Archiving in a repository

A place where things can be stored and shared




# Repositories for data

## 4TU.Centre for Research Data

---

Data underlying the paper: "Fracture Mechanisms and Microstructure in a Medium Mn Quenching and Partitioning Steel Exhibiting Macrosegregation"

 [Celada-Casero, C. \(Carola\)](#)

 [Hidalgo, J. \(Javier\)](#)

 [Santofimia, M.J. \(Maria\)](#)

TU Delft, Faculty of Mechanical, Maritime and Materials Engineering, Department of Materials Science and Engineering

### Digital Object Identifier


<https://doi.org/10.4121/uuid:67e93016-8a24-4381-880c-073975797eac>

# Repositories for data

## 4TU.Centre for Research Data

---

Data underlying the paper: "Fracture Mechanisms and Microstructure in a Medium Mn Quenching and Partitioning Steel Exhibiting Macrosegregation"

 [Celada-Casero, C. \(Carola\)](#)

 [Hidalgo, J. \(Javier\)](#)

 [Santofimia, M.J. \(Maria\)](#)

TU Delft, Faculty of Mechanical, Maritime and Materials Engineering, Department of Materials Science and Engineering

Digital Object Identifier


404 NOT FOUND 

<https://doi.org/10.4121/uuid:67e93016-8a24-4381-880c-073975797eac>

# Repositories for data

## 4TU.Centre for Research Data

Data underlying the paper: "Fracture Mechanisms and Microstructure in a Medium Mn Quenching and Partitioning Steel Exhibiting Macrosegregation"

 [Celada-Casero, C. \(Carola\)](#)

 [Hidalgo, J. \(Javier\)](#)

 [Santofimia, M.J. \(Maria\)](#)

TU Delft, Faculty of Mechanical, Maritime and Materials Engineering, Department of Materials Science and Engineering

### Digital Object Identifier

<https://doi.org/10.4121/uuid:67e93016-8a24-4381-880c-073975797eac>

### How to cite this item

Citation style **Datacite** 

Celada-Casero, C. (Carola); Hidalgo, J. (Javier); Santofimia, M.J. (Maria) (2019) Data underlying the paper: "Fracture Mechanisms and Microstructure in a Medium Mn Quenching and Partitioning Steel Exhibiting Macrosegregation". 4TU.Centre for Research Data. Dataset. <https://doi.org/10.4121/uuid:67e93016-8a24-4381-880c-073975797eac>

# Repositories for software



+

zenodo



**OPEN**  
**RESEARCH SOFTWARE**  
& OPEN SOURCE



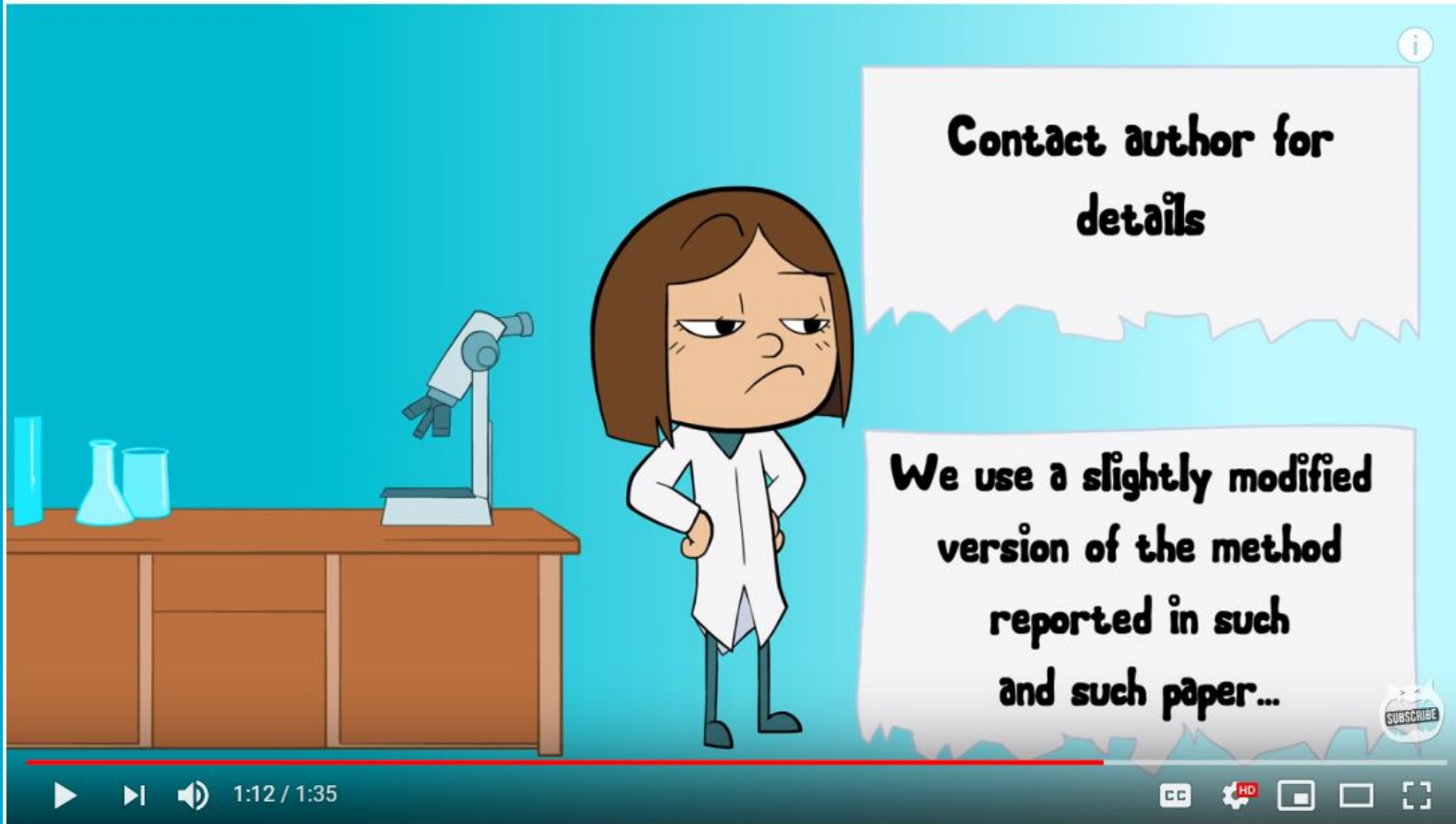
Module 5, Task 2: How to make your code citable using GitHub and Zenodo

277 views

👍 9    💬 0    ➦ SHARE    ≡+ SAVE    ...

<https://www.youtube.com/watch?v=pjsbBQYOOaE&t=1s>  
<https://guides.github.com/activities/citable-code/>

# Repositories for protocols



Protocols.io - Share science protocol knowledge

<https://www.youtube.com/watch?v=84B8P6BAOgM>  
<https://www.protocols.io/>



# Licences for data

[Public Domain Dedication \(CC0\)](#)

Attribution (CC BY)

Attribution-NoDerivatives (CC BY-ND)

Attribution-NonCommercial (CC BY-NC)

Attribution-NonCommercial-ShareAlike (CC BY-NC-SA)

Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND)

# Licences for software and code

MIT License

Apache Licence 2

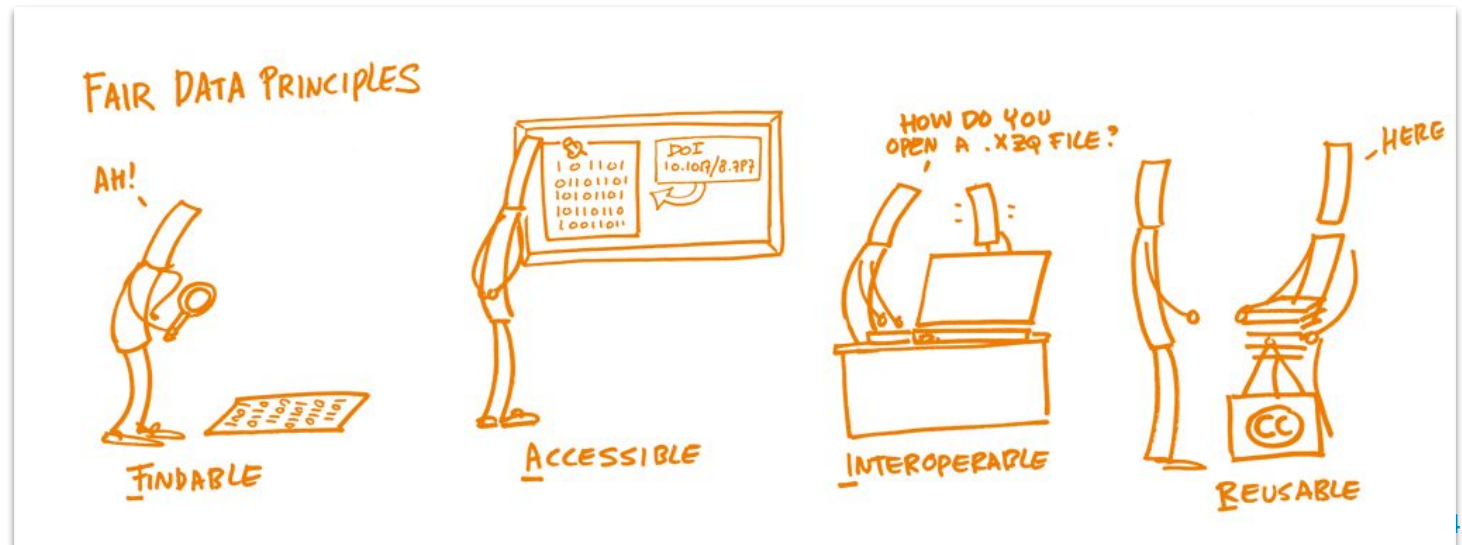
GNU General Public Licence 3 (GNU GPLv3)

<https://researchdata.4tu.nl/en/use-4turesearchdata/archive-research-data/upload-your-data-in-our-data-archive/licencing/>

# Funders' & Publisher requirements



[http://ec.europa.eu/research/press/2016/pdf/opendata-infographic\\_072016.pdf#view=fit&pagemode=none](http://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf#view=fit&pagemode=none)



# Thank you Questions?

