

# Managing your research data

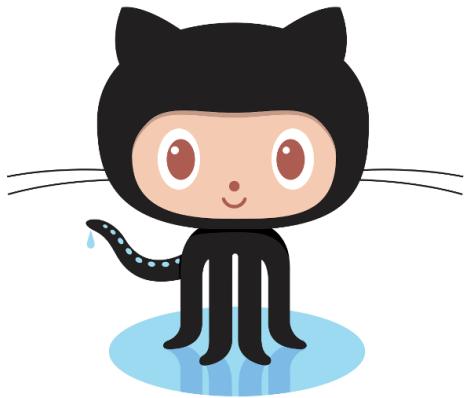


Sergio Martínez Cuesta

# Aims for today

- What is research data?
- Why managing your data can be useful for you and others?
- Challenges
- Gaining confidence to organise your data well
- Existing resources and tools

# Materials available at ...



<https://github.com/semacu/training>  
folder: 20170601\_RDM\_Wolfson/

# What is research data?

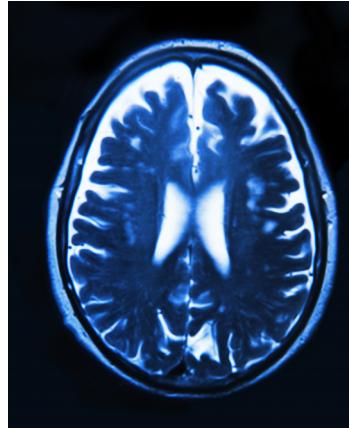
Pieces of information that are descriptive of the research object ... or are the object itself

- ✓ Raw/processed data produced at a research facility
- ✓ Published dataset

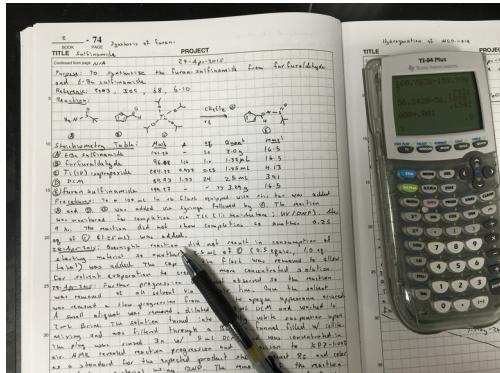
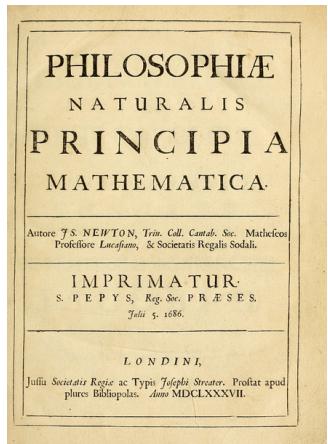
***Necessary to validate research findings***



## Experimental data



## Images



## Primary sources



## Databases



## Email



## Audio



## Questionaries and surveys

... and many others ...

# What is *your* research data?

Talk to the person next to you for 2 min and exchange information:

- Who are you?
- What kind of research do you do?
- What types of data do you collect?



# How to manage your data?

The everlasting external disks



*But are they really permanent? What if ...?*

# Why the need to manage it ?

*“My field is very competitive and I can’t risk **wasting time** with all of this, so I’d rather do **real** research than tidy up my data”*

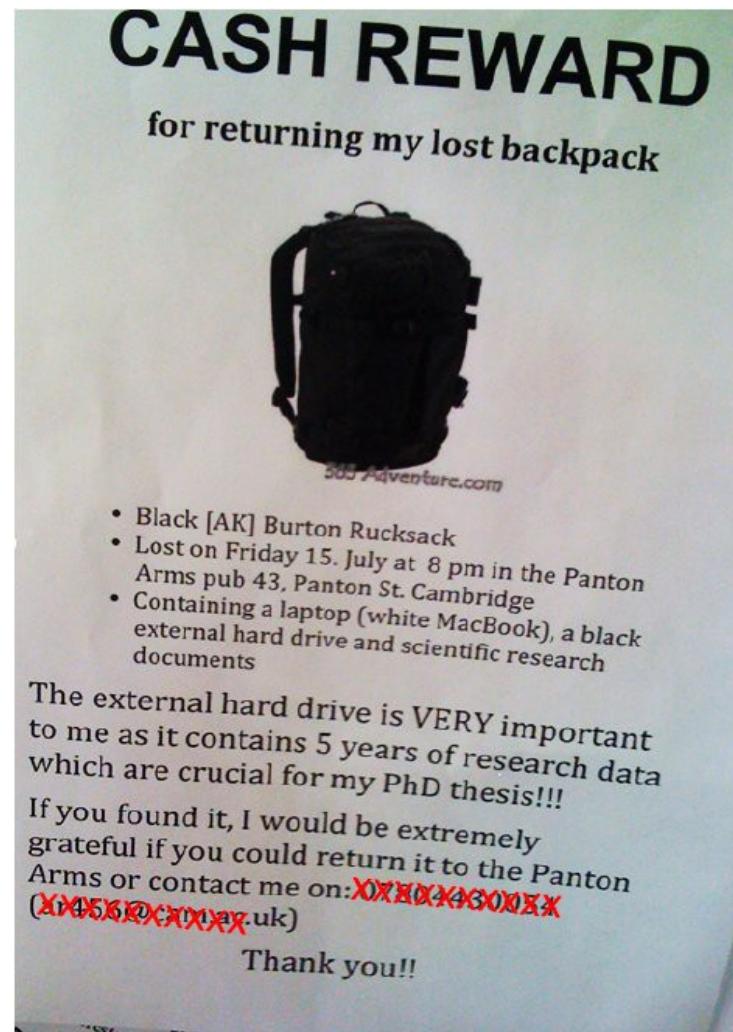
*“My data **are spread** over so many hard drives and directories that it would just be too much work to collect them all in one place”*

*“I can always sort out all my data **after submission** anyway”*

Markowetz F., 2015

# To avoid data *disasters* ...

What would you do if you'd lose your data tomorrow?



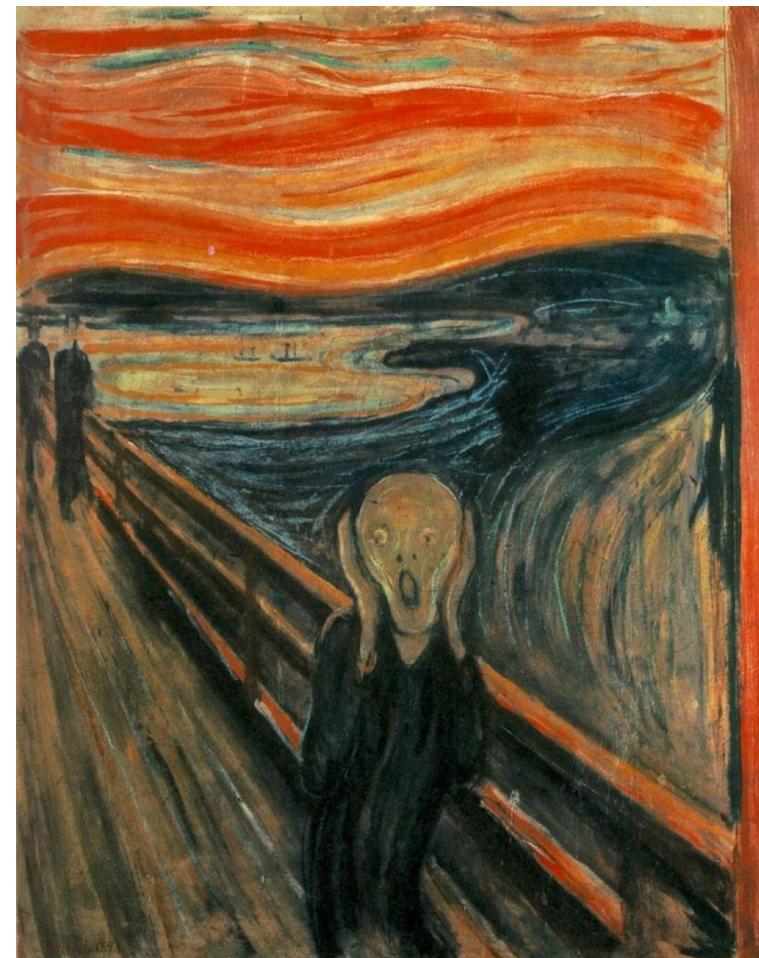
# What would you do if you'd lose your data tomorrow?



*Cancer Research UK – University of Manchester – 27 April 2017*

# What would you do if you'd lose your data tomorrow?

- Your laptop got stolen
- Your office/house burnt
- Your USB stick is lost
- Your portable hard disk is damaged
- Data copied to Dropbox disappeared



# Data backup

*At least 2 backups at 2 different locations*

## *External disks*



## *Online backup*



## *Servers*

Department  
College  
IT



Cheap  
£10-15 / TB (1024GB)



Failure rate  
1.5%/year

Accessibility  
Free (limit)

Personal data  
Hacking

Managed by  
experts

Moving between  
institutions

# Data backup



*Manual*

Copying files

*Automated*

Time machine

Automatic upload  
files as any  
changes are  
saved

Revert to old  
versions

OneDrive 15GB  
free

Dropbox 2GB  
free

Google Drive  
15GB free

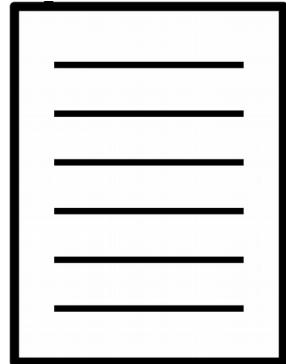
DS-FileStore

# If manual, how often?



## How much would you be willing to lose?

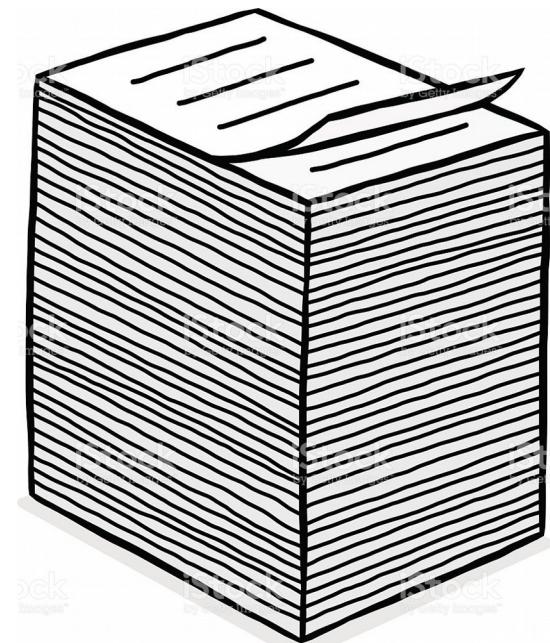
1 day



1 week



1 month-year



# To allow **continuity** of your work ...

*“I obtained the data 6 months ago. I am too busy. Of course **I can’t remember** all the details of all my projects after such a long time”*

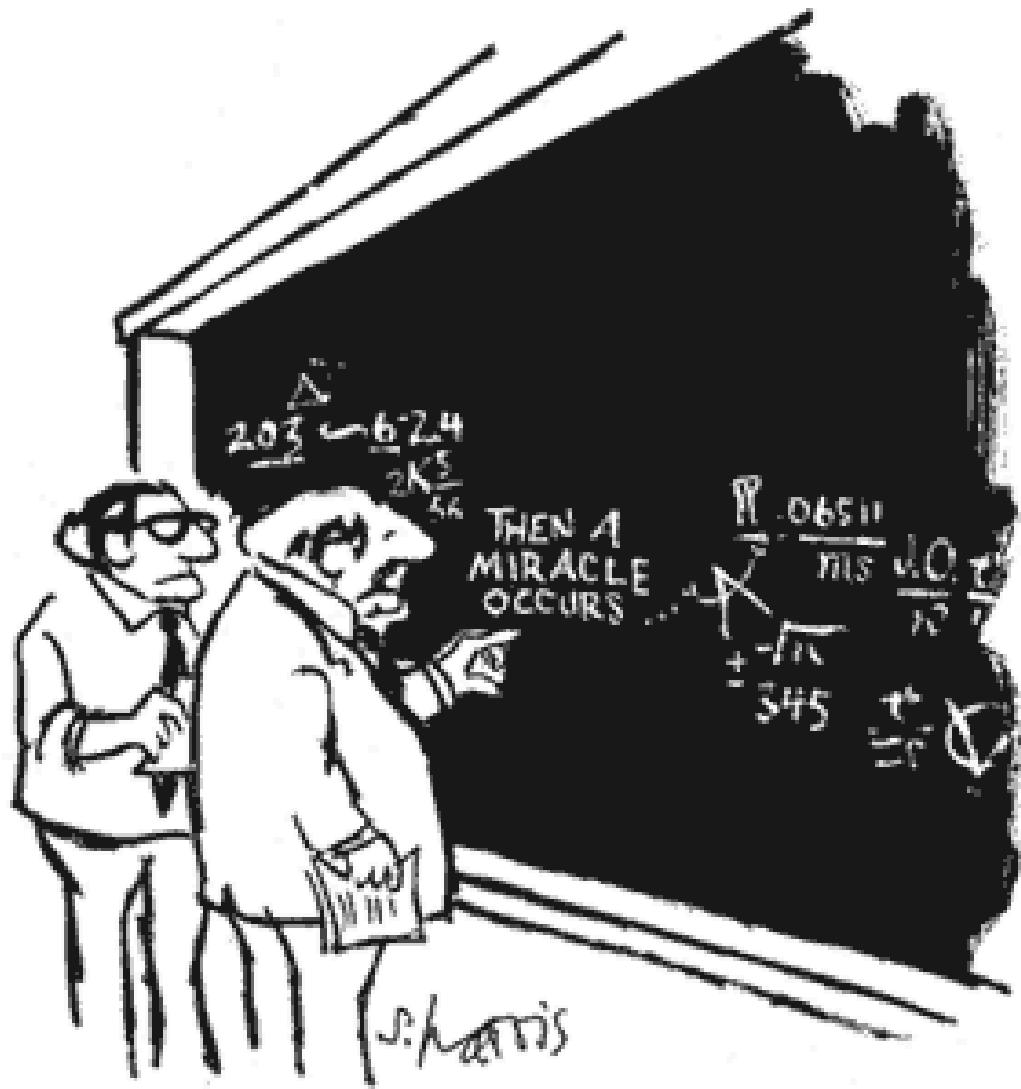
*“My supervisor said I should continue the project of a previous student, but that student is long gone and **hasn’t saved any data or documentation**”*

# Data backup and file sharing

Makes it easier to *write papers* ...

Build your *reputation* ...





"I THINK YOU SHOULD BE MORE  
EXPLICIT HERE IN STEP TWO."

# When do I need to worry about managing my data?

Always!

- Before you start the project
- Before collecting the data
- While you do the analysis
- When you write/co-author a paper
- When you review a paper

# The basics

- Keep your project organised
- Name your files and directories consistently in some informative way
- Start your data at a single backed-up location

# References

Materials

<https://github.com/semacu/training>

Research Data Team

<http://www.data.cam.ac.uk/>

Five selfish reasons to work reproducibly

<https://dx.doi.org/10.1186/s13059-015-0850-7>