

# 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?

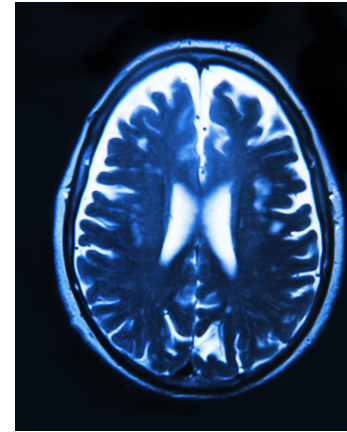
Pieces of information that are descriptive of the research object ... or are the object itself

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

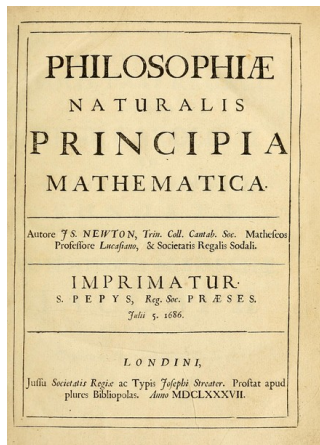
***Necessary to validate research findings***



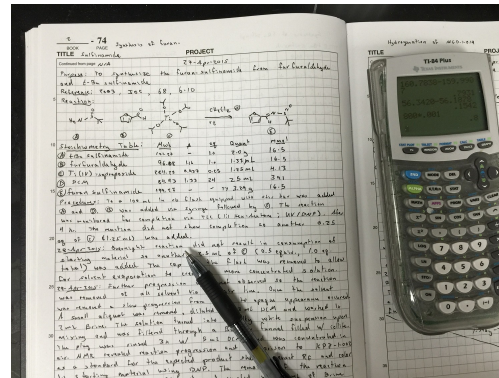
Experimental data



Images



Primary sources



Notebooks and diaries



Questionnaires and surveys



Databases



Email



Audio

... 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”*



# To avoid data *disasters* ...

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

**CASH REWARD**  
for returning my lost backpack



500 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: ~~XXXXXX300XX~~  
~~XXXXXX300XX~~ (XXXXXX300XX@XXXXXX.uk)

Thank you!!

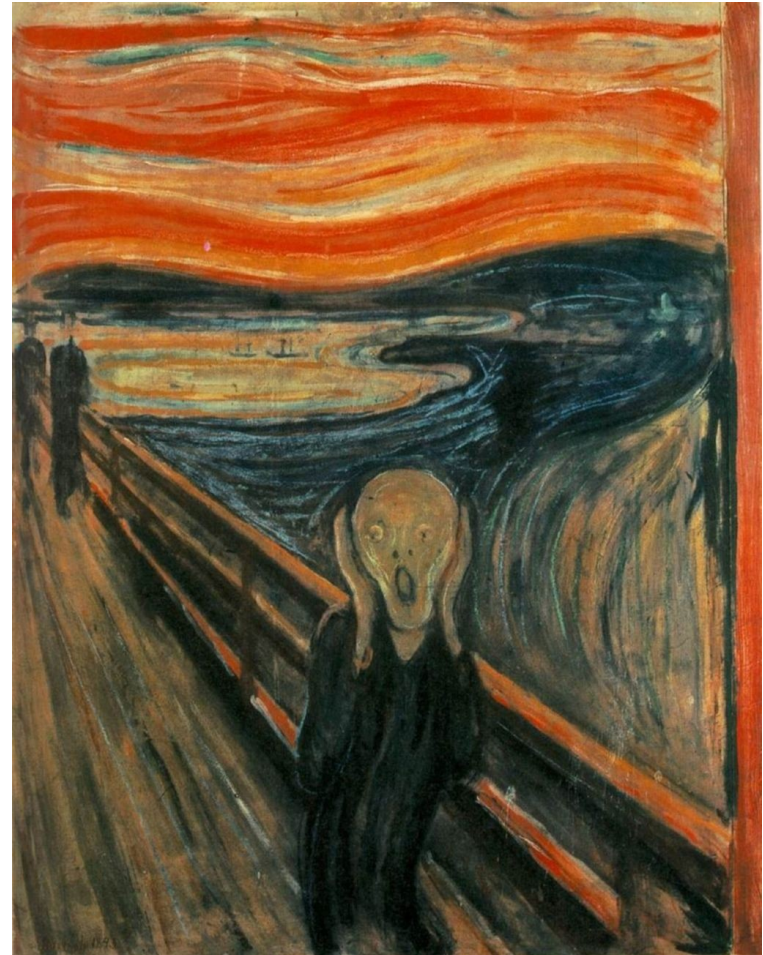
# What would you do if you'd lose your data tomorrow?



*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***



**Cheap**  
£10-15 / TB (1024GB)



**Failure rate**  
1.5%/year

## ***Online backup***



**Accessibility**  
**Free (limit)**

**Personal data**  
**Hacking**

## ***Servers***

Department  
College  
IT



**Managed by**  
**experts**

**Moving between**  
**institutions**



# Data backup

## *Manual*

Copying files to relevant folders



Copying files to relevant folders



## *Automated*

Install software  
e.g. Time machine  
(Mac users)



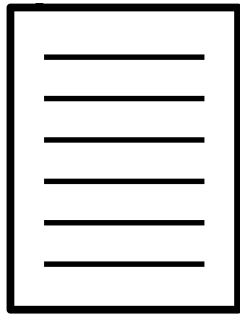
Automatically upload  
files to the cloud when  
any changes are saved

# If manual ... how often?



## How much would you be willing to lose?

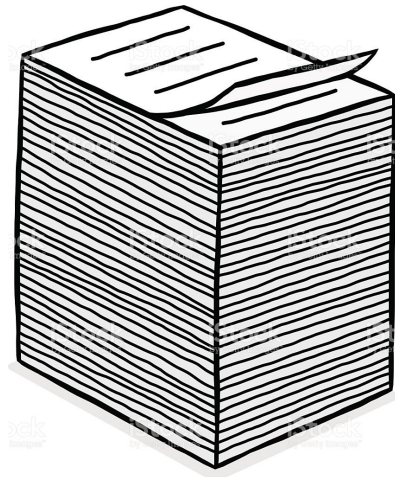
1 day



1 week



1 month-year



*Software allows you to set up **backup time** automatically*

# Data backup and file sharing



<b><i>Space/price</i></b>	2 GB (free) Unlimited (£55/year)	15 GB (free) 1 TB (~£80/year)	1 TB (free)
<b><i>File history and recovery</i></b>	Yes, unlimited	Yes	Last 90 days
<b><i>File size limit</i></b>	None	5 GB	15 GB
<b><i>Support</i></b>	UIS	Unsupported	UIS
<b><i>OS</i></b>	Windows, Mac, Linux, Android, iOS	Windows, Mac, Android, iOS	Windows, Mac, Android, iOS
<b><i>Accessibility</i></b>	Sync anywhere on any devices	Live editing	Integration with Microsoft Office

# More ... file sharing



Email



Website



FTP



# How do you **manage** your research data?

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

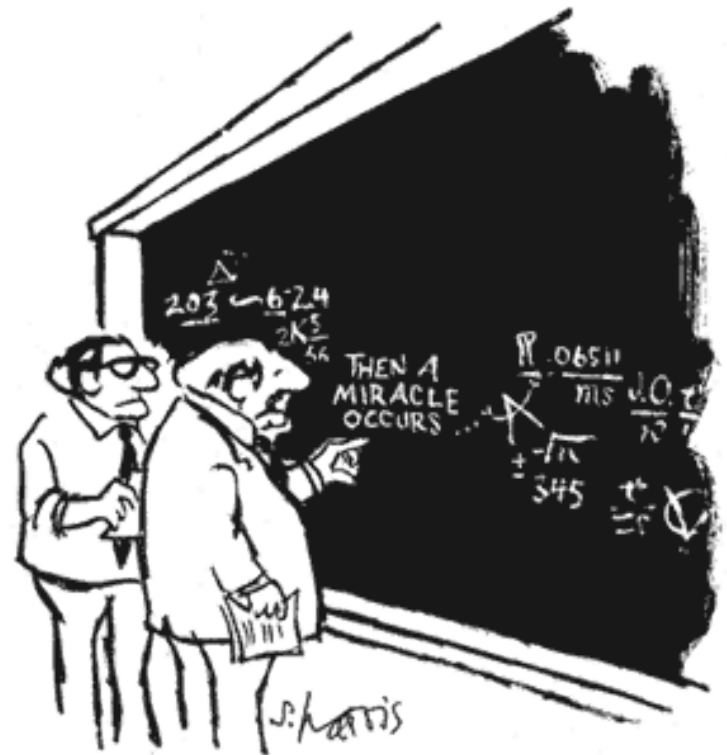
- Do you do any data **backup**?
- How **often**?
- How do you **share** files/data with collaborators?



# *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"*

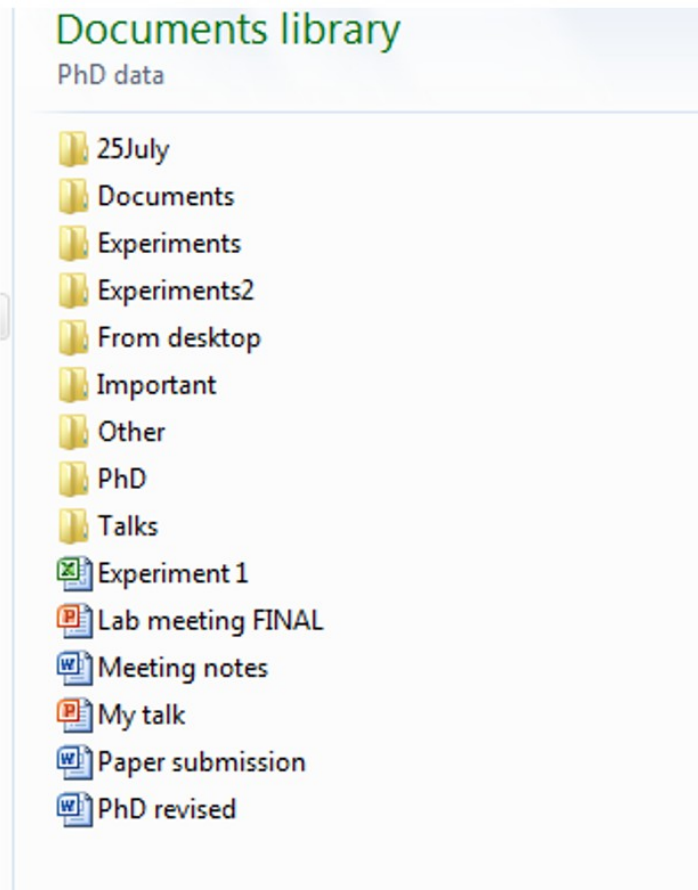
*"My supervisor said I should continue the project of a previous student, but that student is long gone and **hasn't saved any data or documentation**"*



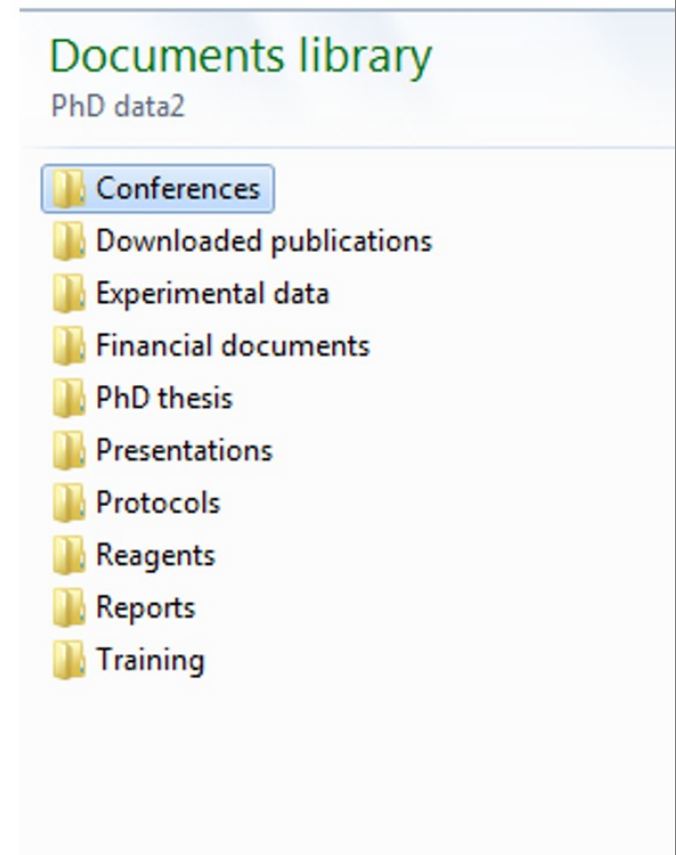
"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

# Data organisation

## Example A



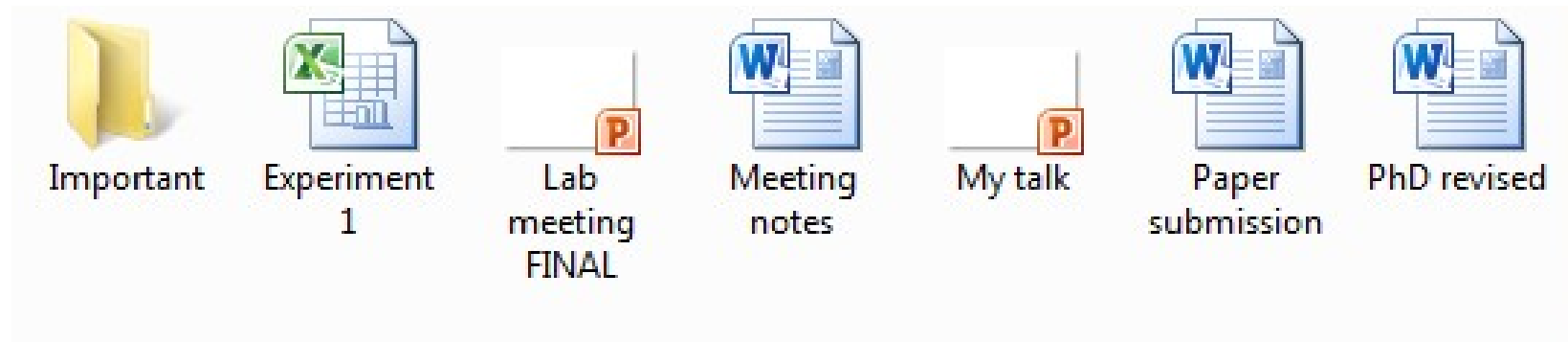
## Example B



# Data organisation

- Consistent
- Meaningful to you and your colleagues
- Allow you to find files easily

# File naming



*Create a naming convention that works for you:*

fk468\_PCF-b2-oxhyd-1.fastq.gz

Collaborator: fk (Fumiko Kawasaki)

Experiment number: 468

Life cycle stage: PCF

Batch: b2

Treatment: oxhyd

Replicate: 1


















.../20170601\_RDM\_Wolfson/...

Date: 20170601

Topic: RDM  
(Research Data Management)

Location: Wolfson college

# File naming

- ▶  20160706\_Presentations
- ▶  20160726\_RDMworkshopForLibrarians\_MillLane
- ▶  20160802\_Altmetrics
- ▶  20160804\_Bibliometrics
- ▼  20160810\_RDMworkshopForLibrarians\_MilsteinRoom
  -  GDL\_DMP\_V8\_2016511.docx
  -  GDL\_DMPForLibrarians\_V9\_20160726docx.docx
  -  GDL\_ExampleDMP\_V9\_20160511.docx
  -  LST\_RDMforLibsAttendees\_V1\_20160808.pdf
  -  MEM\_RDMforLibs2Feedback\_V1\_20160811.docx
  -  PRE\_DataLossScenarios\_V5\_20160726.docx
  -  PRE\_RDMforLibsSigns\_V1\_20160808.pptx
  -  PRE\_RDMWorkshopForLibrarians\_V3\_20160726.pptx
  -  PRE\_RDMWorkshopForLibrarians\_V4\_20160810.pdf
  -  PRE\_RDMWorkshopForLibrarians\_V4\_20160810.pptx
- ▶  20160914\_RDMWorkshopforGLS
- ▶  20161004\_PhDTraining

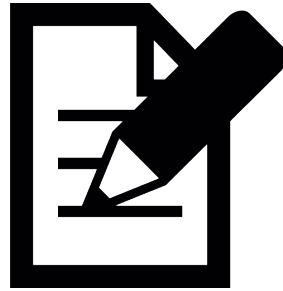
# How do **organise** your research data?

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

- How do you organise your files?
- How do you name your files?



Easier to ***write papers*** ...



Build your ***reputation*** ...

The FAIR principles:



**Findability**  
**Accessibility**  
**Interoperability**  
**Reusability**



# 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 writing/co-authoring your paper/thesis

When reviewing the work of others

# Overview

Try to keep your projects ***organised***

Name files and directories ***consistently*** using some ***informative*** way

Store your data at a ***single backed-up*** location

# Questions?



## 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>