

**STUDENTS VERSION**

**UNIVERSITY  
OF TWENTE.**

**RESEARCH DATA MANAGEMENT**

**2021 VERSION 1.1**

**JUDITH BRANDS**



# Research Data Management



Active organization and maintenance of data throughout the research process

# Why research data management (RDM)?

There are many reasons why research data management is important.

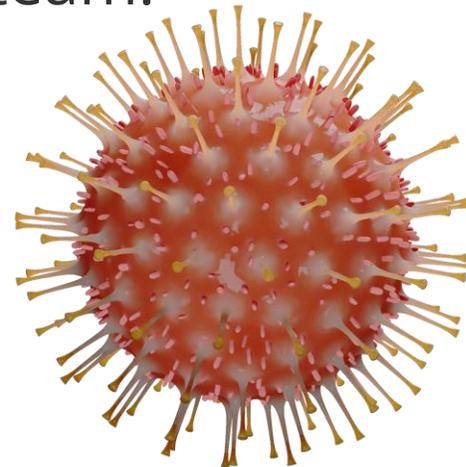
1. Good scientific practice.
2. Growing research data requirements imposed by funders and publishers.
3. Research data management saves time and resources in the long run.
4. Good management helps to prevent errors and increases the quality of your analyses.



# Why research data management (RDM)?

There are many reasons why research data management is important.

5. Well-managed and accessible data allows others to validate and replicate findings.
6. Research data management facilitates sharing of research data and, when shared, data can lead to valuable discoveries by others outside of the original research team.



Covid-19





Sure sure....but I don't understand what it means for me....

Why am I here?

# Why research data management (RDM)?

There are many reasons why research data management is important for you!

Good scientific practice is an important academic skill:

1. It helps you to safely store your data, so nothing gets lost
2. Prevents unauthorized access to your data
3. Helps you to keep track of the work you have done so far
4. Saves you time finding your data
5. Improves the quality of your work and decreases the chance of errors



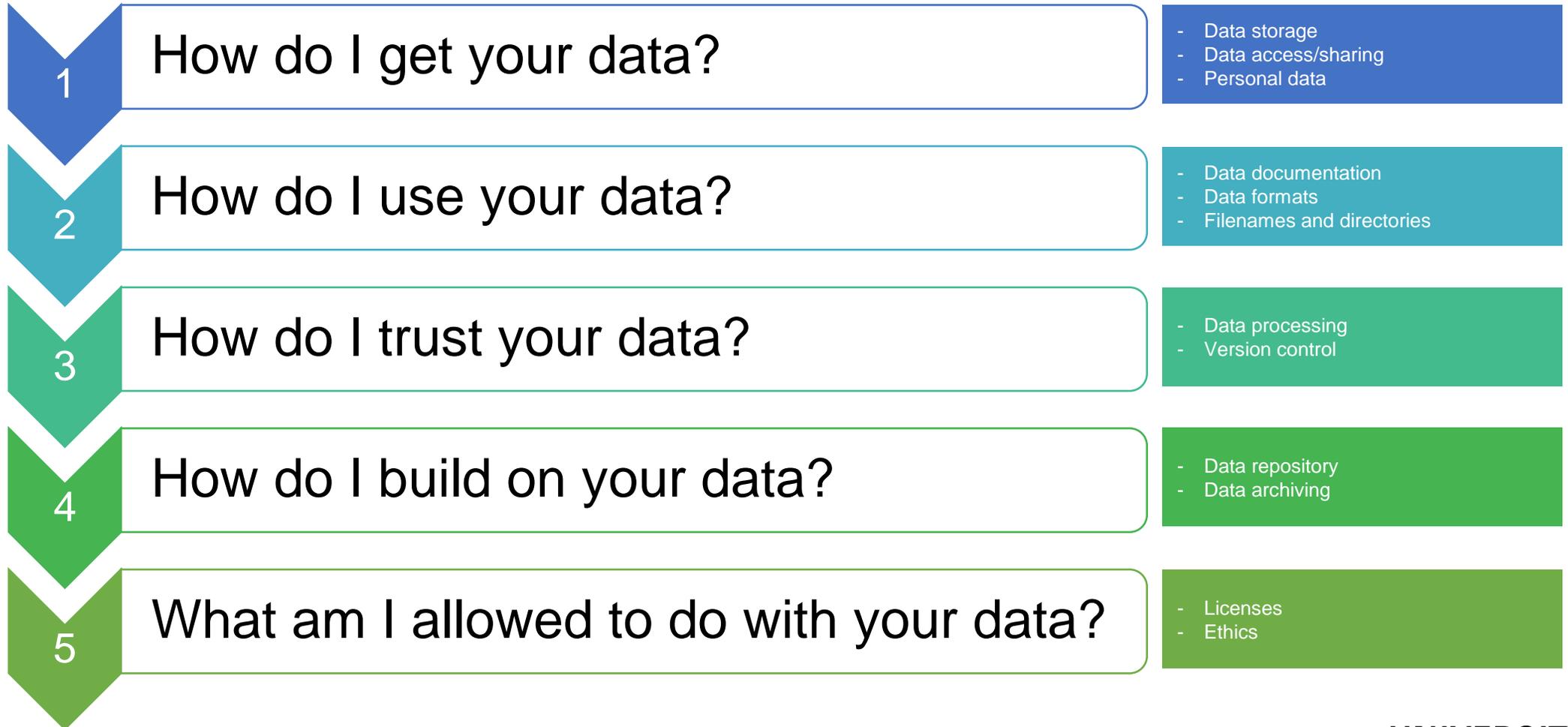
# Be aware....

Let's start with a small disclaimer....Please always check if your group or department has specific guidelines on data management.

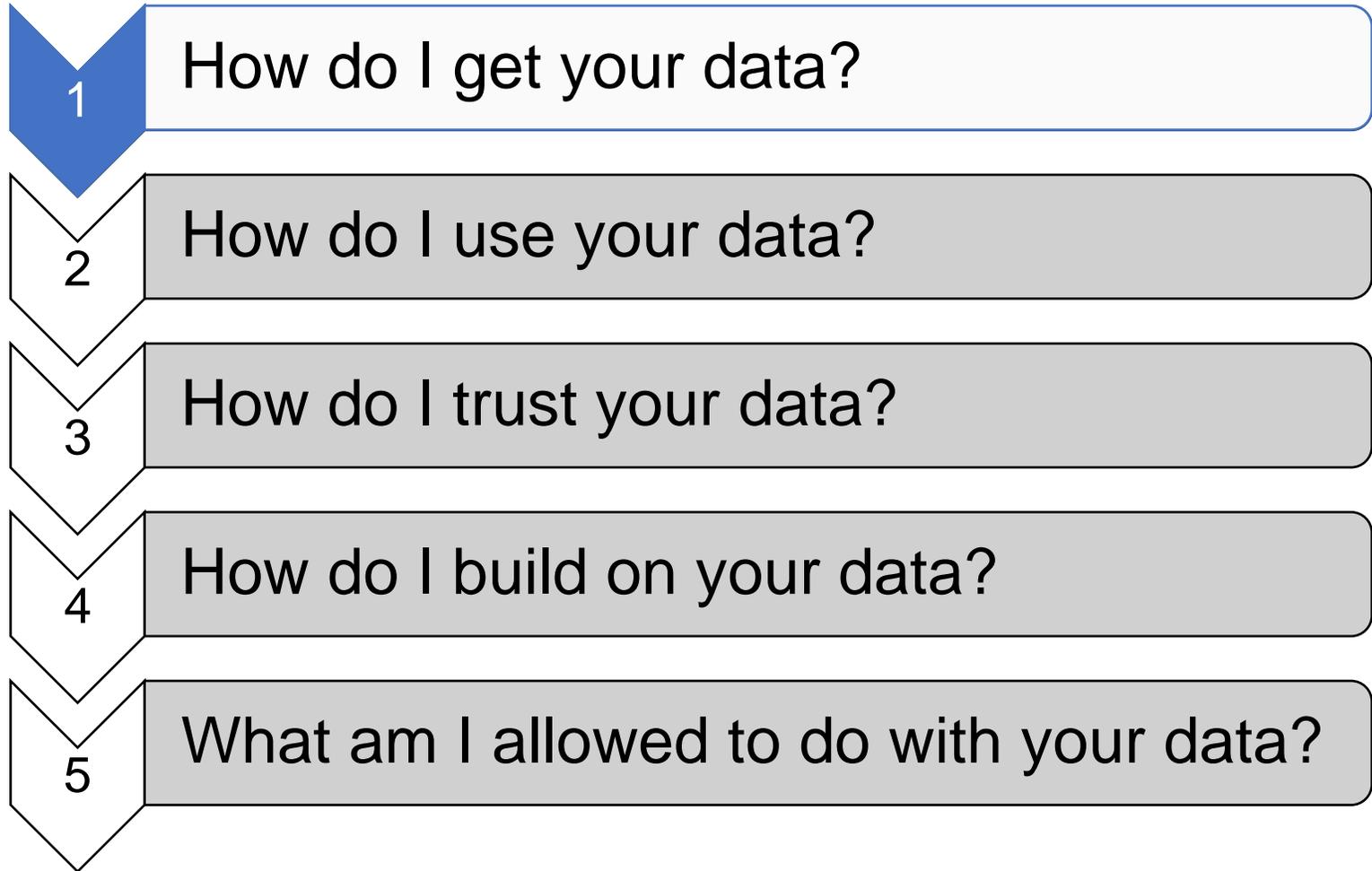
*You can check the [UT data policy](#), your [faculty policy](#) or group guidelines to start with.*



# What will be discussed?



# What will be discussed?

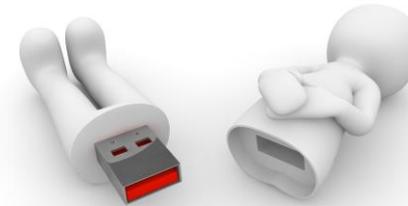


- Data storage
- Data access/sharing
- Personal data

# Data storage

A care provider from the Flevoziekenhuis in Almere has taken the data of more than 4300 patients with them from the hospital. The data was put on a USB stick, but the employee lost it outside in the parking lot.

So perhaps not every type of storage device should be used....but what options do I have?



# Data storage

During your project you have collected some data, where do you store this data?

What about your laptop?



## Data storage

During your project you have collected some data, where do you store this data?

What about your laptop?

Your laptop can die, get lost or stolen. When this happens you have lost all your data. Besides, how will I have access to your data?



# Data storage

During your project you have collected some data, where do you store this data?

What about an external hard drive?



## Data storage

During your project you have collected some data, where do you store this data?

What about an external hard drive?

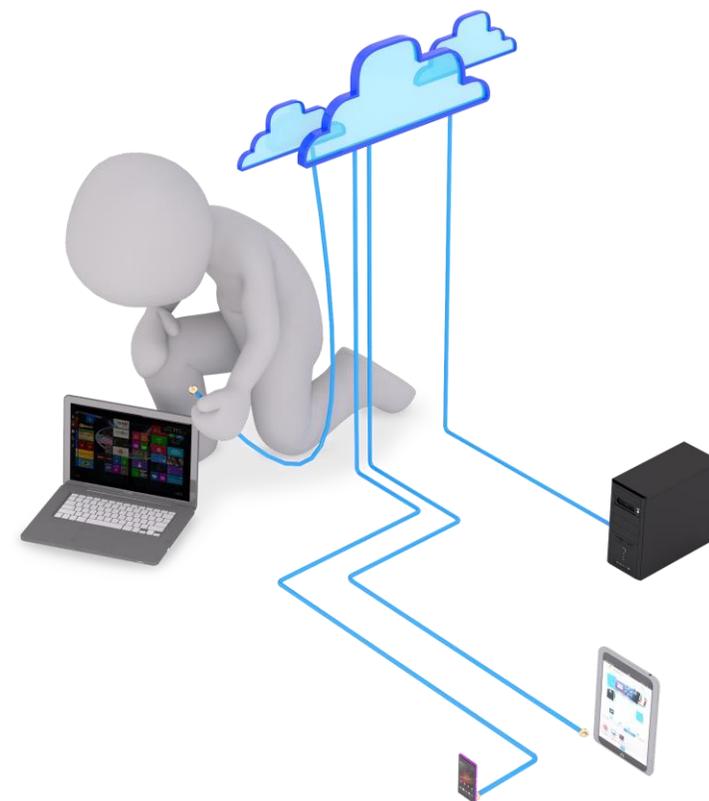
It is not uncommon for external hard drives to quit working or get corrupted. As shown, USB sticks and external hard drives can get lost or stolen. Another risk for losing all your hard work. In addition, I still don't have access to your data.



# Data storage

During your project you have collected some data, where do you store this data?

Okay, so what about cloud services?



# Data storage

During your project you have collected some data, where do you store this data?

Okay, so what about cloud services?

Great idea! But you can't use just any cloud service. If you use your personal account, there will be no certification and the data is usually stored on a server outside the EU. This is not very secure, not quite good enough. Nonetheless, I like where this is going.



## Data storage

During your project you have collected some data, where do you store this data?

What if I use my UT account with the cloud services?

# UNIVERSITY OF TWENTE.

# Data storage

During your project you have collected some data, where do you store this data?

What if I use my UT account with the cloud services?

There are agreements between the university and certain cloud services, as a result these are now certified storage solutions. Data stored in the cloud can be shared, so I can finally have access to your data.



# Data storage

During your project you have collected some data, where do you store this data?

What if I use my UT account with the cloud services?

You can use this decision tree to get some more information

<https://webapps.utwente.nl/dmp/fr/RDM/Storage-Decision-Tree/new>



# Data storage

During your project you have collected some data, where do you store this data?

What if I use my UT account with the cloud services?



# Data storage

During your project you have collected some data, where do you store this data?

Are there any other options?



# Data storage

During your project you have collected some data, where do you store this data?

Are there any other options?

There is the Project-Drive. Your group or mentor will have space here where you can store your data. This is a certified server. Data is stored locally at the UT and data can be shared, even with people outside the university.



## Let's summarize

During your project you have collected some data, where do you store this data?



Use cloud services with your UT account



Use the P-Drive using the account of an employee

## Let's summarize

During your project you have collected some data, where do you store this data?



Use cloud services with your UT account



Use the P-Drive using the account of an employee



Don't store data on your personal laptop.



Don't use external hard drives or USB sticks for data storage

## Let's summarize

During your project you have collected some data, where do you store this data?

**Never** store personal identifiable data of your research participants on your personal laptop or an unprotected USB-device! Those data need sufficient protection under GDPR (dutch AVG). Later more on this topic



Don't store data on your personal laptop.



Don't use external hard drives or USB sticks for data storage

## Let's summarize

During your project you have collected some data, where do you store this data?



Use cloud services with your UT account



Use the P-Drive using the account of an employee



Don't store data on your personal laptop.



Don't use external hard drives or USB sticks for data storage



Don't use a cloud service with a personal account

## Data storage

During your project you have collected some data, where do you store this data?

WAIT!!

Why would you need access to my data?



# Data sharing/access to data

Why is data sharing important?

-  What if I'm a colleague working with you on a project?
-  I could follow-up on your project after you have finished....
-  It's for researchers important their supervisor/mentor/PI has access to the data in case something happens (e.g. sickness), so the project isn't on hold or stopped completely. By now we know anything can happen....

# Data sharing/access to data

Should everybody be able to access my data?

# Data sharing/access to data

Should everybody be able to access my data?



## Data sharing/access to data

Should everybody be able to access my data?

A secretary from a hospital in Roosendaal has looked at the medical files of acquaintances - including her husband's ex-wife - for years without the hospital noticing. The secretary had no treatment relationship with these patients. Medical details then ended up in a revenge novel that the man wrote about his ex-wife.



# Data sharing/access to data

Should everybody be able to access my data?

When your data involves

- personal data such as patient/medical information or certain personal data from, for instance interviews, and informed consent ([AVG/GDPR](#))
- commercial interest / data for patents

you should be very careful who is allowed to have access to your data. Failing to comply can have serious consequences.



# Data sharing/access to data

Should everybody be able to access my data?

When your data involves

- personal data such as patient/medical information or certain personal data from, for instance interviews, and informed consent ([AVG/GDPR](#))
- commercial interest / data for patents

you should be very careful who is allowed to have access to your data. Failing to comply can have serious consequences.



# Data sharing/access to data

Should everybody be able to access my data?

Always check your type of data and who can and should have access to your data.



# Personal data/information/details

So what does this include exactly?

De [Autoriteit persoonsgegevens](#) meld de volgende informatie:

De Algemene verordening gegevensbescherming (AVG) geeft aan dat een persoonsgegeven alle informatie is over een geïdentificeerde of identificeerbare natuurlijke persoon.



The [Autoriteit persoonsgegevens](#) mentions:

The General Data Protection Regulation (GDPR) states that personal data is all information about an identified or identifiable natural person.



# Personal data/information/details

So what does this include exactly?

Dit betekent dat informatie ofwel direct over iemand gaat, ofwel naar deze persoon te herleiden is. Er zijn veel soorten persoonsgegevens. Voor de hand liggende gegevens zijn iemands naam, adres en woonplaats. Maar ook telefoonnummers en postcodes met huisnummers zijn persoonsgegevens.

This means that information is either directly about someone or can be traced back to this person. There are many types of personal data. Obvious data is a person's name, address and place of residence. Telephone numbers and postal codes with house numbers are also personal data.

# Personal data/information/details

There is more....

Gevoelige gegevens als iemands ras, godsdienst of biometrische en genetische informatie worden bijzondere persoonsgegevens genoemd. Deze zijn door de wetgever extra beschermd.

Het is verboden om bijzondere persoonsgegevens te verwerken, tenzij er een wettelijke uitzondering is.

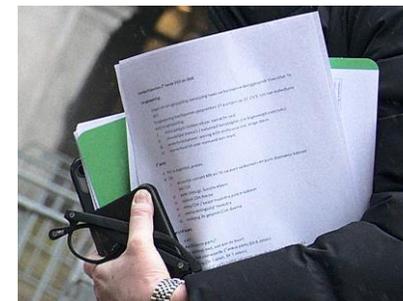
Sensitive data such as a person's race, religion or health are referred to as special personal data. These are extra protected by the legislator.

It is prohibited to process special personal data, unless there is a legal exception.

# Sensitive data storage

**Any** sensitive data need to be stored extra carefully to prevent a data breach.

- No laptops
- No external hard drives (when possible)
- Don't walk around with pieces of paper with sensitive data
- Don't use (random) cloud services (e.g. dropbox or using your personal account)



Exploratory research  
regarding the formation of  
the government

# Sensitive data storage

**Any** sensitive data need to be stored extra carefully to prevent a data breach.

When it goes wrong?



# When it went wrong

## Red Cross apologises after mass leak of Australian blood donor records



**Paul Smith**  
Technology editor

Updated Oct 28, 2016 - 1:54pm,  
first published at 12:49pm



Save



Share

The Australian Red Cross Blood Service has issued an apology after personal details of 550,000 blood donors were leaked online in one of the country's largest ever data breaches.

The data breach was discovered on Tuesday when an [online service set up to register vulnerable or hacked websites](#) was contacted by a concerned user, who had come across the trove of private data, and the Red Cross confirmed the problem at lunchtime on Friday.

The details contained in a publicly accessible folder included 1,286,366 records, with 550,000 related to identified individuals. It contained information including name, address, contact information, blood type and donation history.

# When it went wrong



**Trouw**



VOORPAGINA

VERDIEPING

OPINIE

RELIGIE&FILOSOFIE

DUURZAAMHEID&NATUUR

CULTUUR&MEDIA

ACHTERPAGINA

Callcenter

## **Autoriteit Persoonsgegevens onderzoekt datalek bij GGD-testlijn. Wie kon privacygevoelige informatie inzien?**



Een callcenter van de GGD in Leiderdorp. Beeld Photo News

# When it went wrong



## Datalek bij Donorregister betreft gegevens van zes miljoen Nederlanders

Alert  01-05-2020



### Ook gedupeerd?

 Neem contact op met de redactie

### Dossier Datalekken 160 items

-  Gegevens van negen miljoen easyjet-klienten gestolen door hackers
  -  Datalek bij Donorregister betreft gegevens van zes miljoen Nederlanders
  -  Ferd Grapperhaus adviseert burgers om NL-Alert app te verwijderen wegens datalek
  -  Datalek bij Nintendo: 160.000 Nintendo Network ID's gehackt
  -  Porno-afpersmail met jouw échte wachtwoord in omloop
-  **NAAR COMPLETE DOSSIER**

# When it went wrong

Home Corona Over privacy ▾ Onderwerpen ▾ Zelf doen ▾ Publicaties ▾ Contact ▾ Info voor FG's

## Boete gemeente Enschede om wifitracking

Persbericht / 29 april 2021

Categorie:

Internet en telecom, Internet of things,  
Overheid & de AVG, Gemeenten

De Autoriteit Persoonsgegevens (AP) geeft de gemeente Enschede een boete van 600.000 euro, omdat de gemeente wifitracking gebruikte in de binnenstad op een manier die niet mag. Daardoor was het mogelijk winkeland publiek en mensen die in de binnenstad wonen of werken te volgen.

### Wifitracking in Enschede

De gemeente Enschede besloot in 2017 om via sensoren de drukte in de binnenstad te gaan meten. De gemeente huurde daarvoor een bedrijf in dat is gespecialiseerd in het tellen van passanten.

Meetkastjes in de winkelstraten ving de wifisignalen op van de mobiele telefoons van passerende mensen. Ieders telefoon werd apart geregistreerd, met een unieke code.

Door te tellen hoeveel telefoons er op een bepaald moment rond een meetkastje zijn, weet je hoe druk het is. Houd je over een langere periode bij welke telefoon langs welk meetkastje komt, dan verandert dit 'tellen' in

### Publicaties

Rapport / 29 april 2021



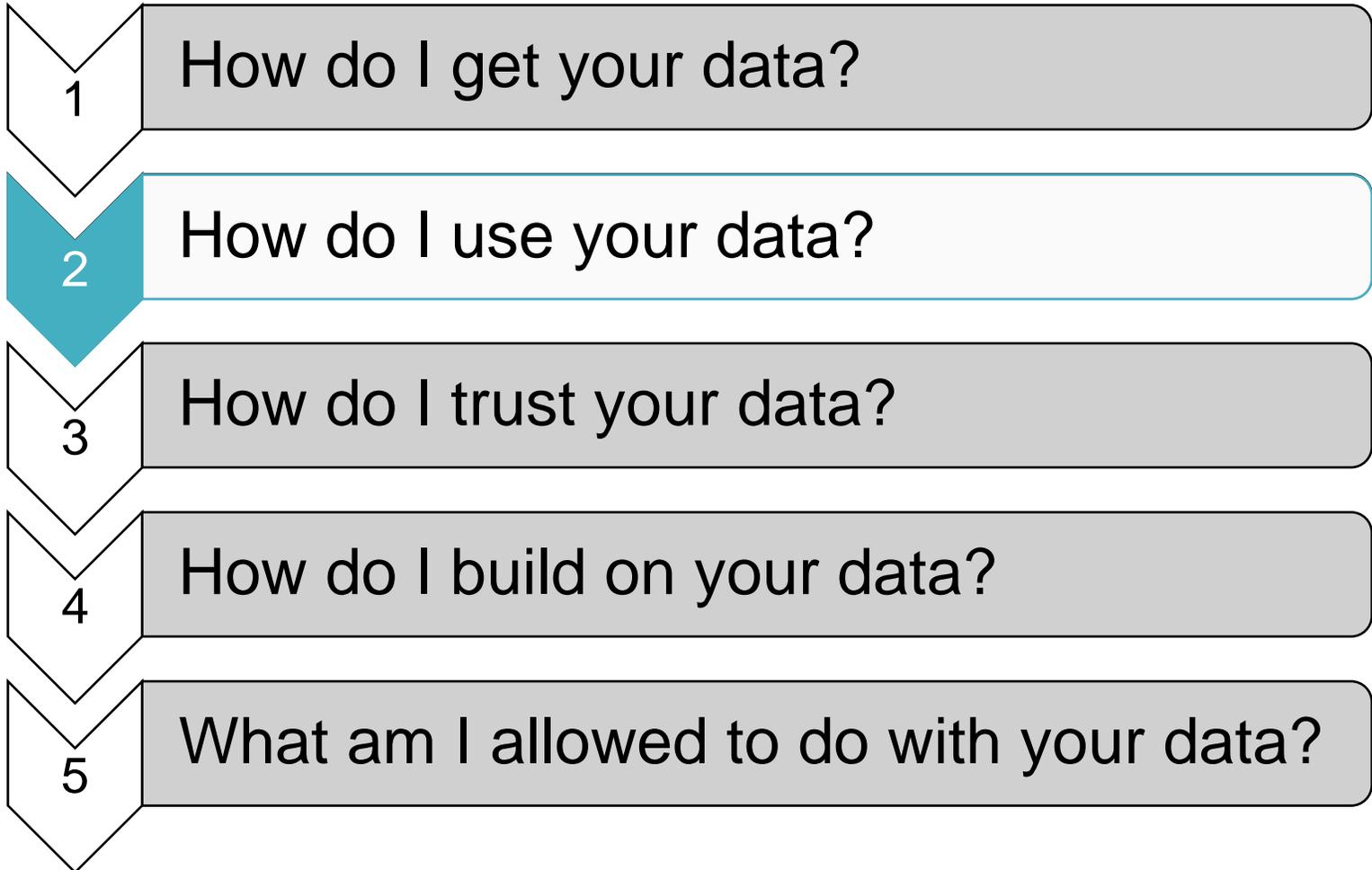
Boete wifitracking  
Enschede

↓ DOWNLOADEN

### Informatie over wifitracking

Veelgestelde vragen van gemeenten en bedrijven over wifitracking

# What will be discussed?



- Data documentation
- Data formats
- Filenames and directories

# Data documentation

Great, I now have access to your data. Will I be able to understand what you have done? Where can I find what part of the data? How was it analyzed?

Have you documented your data?



# Data documentation

Let's turn the tables for just a second....if you would pick up a project, what would you need?

*Any information you need to continue with someone else's project, is information you yourself have to document and store so others can work with your data.*



# Data documentation

Protocols → A protocol outlines the methods, tools, and materials to complete a procedure. They should have enough information to replicate the study.

*Have you ever tried backing a cake without a recipe (like in a great bake-off show?)....it's difficult and many times it just can't be done. A protocol is like a recipe, a step by step description of how to use the materials to get a product.*





# Data documentation

Lab notebook → A scientific diary, documenting your thoughts, ideas, meetings, and notes. Some examples of what you need to include:

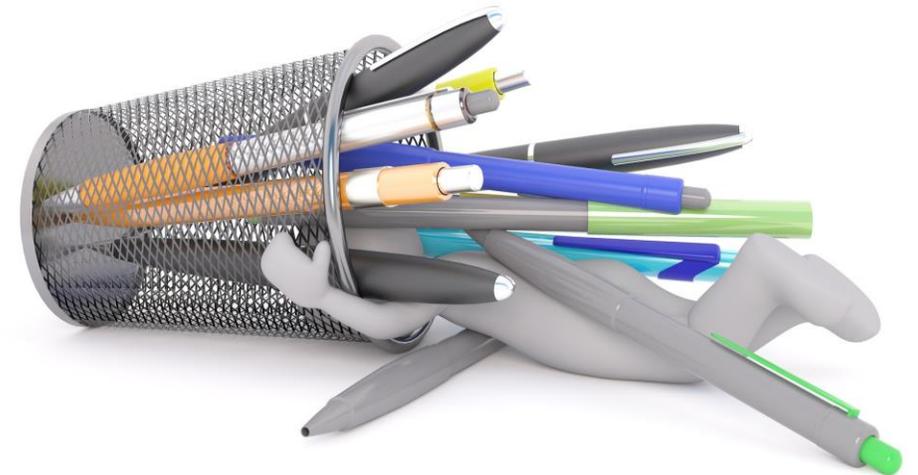
- What were you trying to do?
- How were you doing it?
- Settings of your variables
- Interpretation (failed/worked)
- What do you do with outliers?
- Unexpected results
- Your comments and assumptions
- Locations of samples, files etc.
- Notes from meetings



# Data documentation

Lab notebook → A scientific diary, documenting your thoughts, ideas, meetings, and notes.

It will help navigate yourself and others through your project. What you did, why you did it, why you made changes, what worked, and what didn't. It is the story behind the results.



# Data documentation

Other types of information you need to carefully document



Equipment settings &  
instrument calibration

Questionnaires

Data dictionaries

# Data documentation

Everything you can think of you might need from someone else to take over or continue with their project is information you should document yourself.



# Coffee

So far so good....a little bit more on this topic



Ready to continue?



# Data formats

When you save data you can usually chose a file format. Not every format can be opened by a different operating system. Not everyone might have the same software.

Oh, did you know that file formats might change in time? Microsoft office included, see if you can open my old homework assignment.....

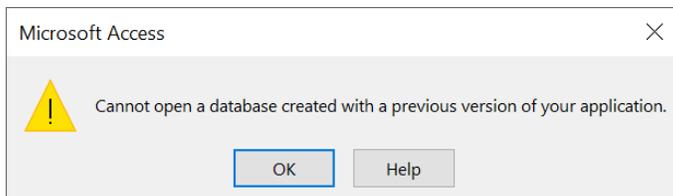


Microsoft Access  
Database

# Data formats

When you save data you can usually chose a file format. Not every format can be opened by a different operating system. Not everyone might have the same software.

Oh, did you know that file formats might change in time? Microsoft office included, see if you can open my old homework assignment.....



You see, I can't.  
Microsoft access has evolved and now I can't open the file anymore.

# Data formats

When possible select a format that is non-proprietary, in other words, independent of the program that created the file.

Sometimes you can do so while working on the data, in other cases you need to convert the files when you are done with the study and ready to 'archive' the data.

# Data formats

When you store your data in a trusted repository preferred formats are used. To learn more about these non-proprietary or preferred file formats, you can check the following documents.

<https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing/file-formats>

 4TU.ResearchData  
UNIVERSITY OF TWENTE

### Preferred File Formats

The choice of file format is of essential importance in order to ensure that the research data will remain usable in the future. 4TU.ResearchData therefore strongly encourages the use of standard, exchangeable or open file formats. For the preferred formats, 4TU.ResearchData guarantees that the research data will remain accessible and that they will be migrated or converted if necessary.

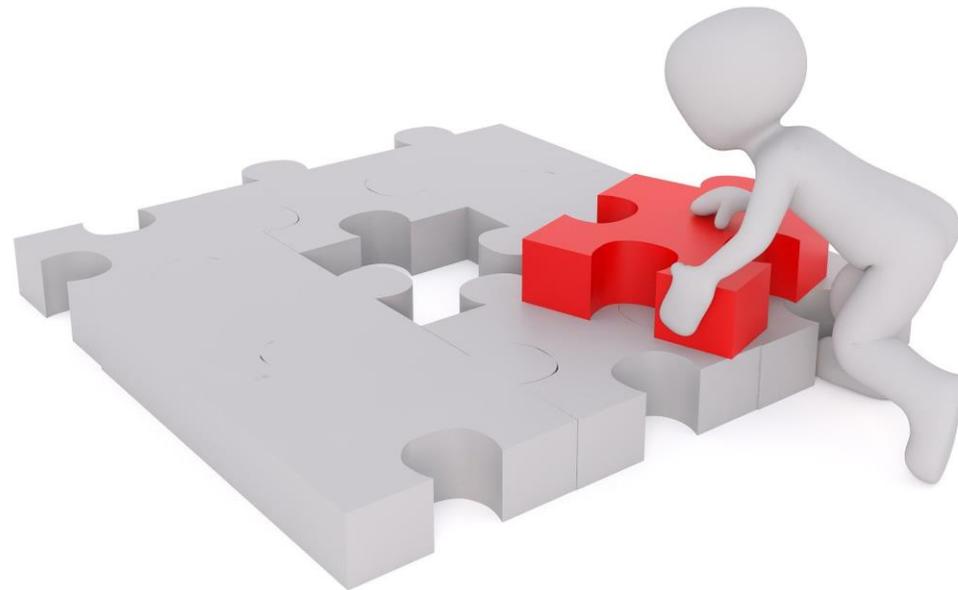
Preferred (sustainable) file formats that receive full preservation	
<b>Text</b>	Plain text, XML, HTML, PDF (PDF/A-1), JSON, PDB (Protein Data Bank), XYZ <i>(all formats should be encoded in UTF-8)</i>
<b>Spreadsheets</b>	CSV (Comma-separated values), Tab-delimited values.
<b>Images</b>	JPEG, TIFF, PNG, SVG
<b>Geospatial</b>	GML (Geographical Mark-up Language), KML (Keyhole Mark-up Language), ESRI Shapefile, Geo-referenced TIFF
<b>Numerical</b>	NetCDF, CSV, JSON
<b>Video</b>	<i>No sustainable format established</i>
<b>Audio</b>	Waveform Audio File Format (WAVE)
<b>Databases</b>	Delimited Flat File w/DDL
<b>Archives</b>	ZIP, TAR, GZIP, 7Z

4TU.ResearchData, Preferred File Formats | September 2019

[4TU.ResearchData's preferred file formats](#)

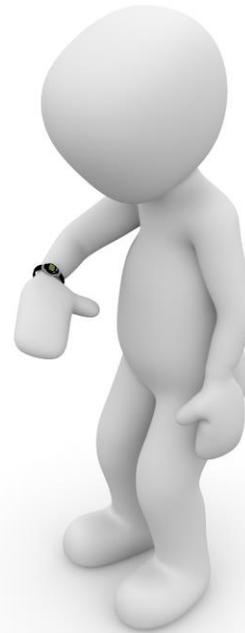
# Filenames and directories

One more thing before we move on; last piece of this puzzle



# Filenames and directories

I collected data for one of my studies. Unfortunately, I couldn't finish the study during this project, and the study was on hold during my next position. By the time I wrote my paper, it was 6 years since the start of data collection! This is not that uncommon in science....



# Filenames and directories

Think about how you want to name your files and directories. Is it clear for you (even in 5 years) and for others what the file is about?

Let's try this. You have done an experiment and save the data. **Exp\_redone.xlsx**

# Filenames and directories

Think about how you want to name your files and directories. Is it clear for you (even in 5 years) and for others what the file is about?

Let's try this. You have done an experiment and save the data. **Exp\_redone.xlsx**



3 months go by (not even years).  
Do you know what is in this file or  
do you need to open it to find out?  
Do you have to check every file  
before you find the one you need?  
That would be quite time  
consuming....

# Filenames and directories

Think about how you want to name your files and directories. Is it clear for you (even in 5 years) and for others what the file is about?

For instance, consider:

Yyyymmdd\_TypeExp\_StudyID\_ResearchID\_Version

Think wisely about the name of your directories as well, it is so easy to get lost and forget where your data is stored. A good plan/structure saves you very precious time, trust me!



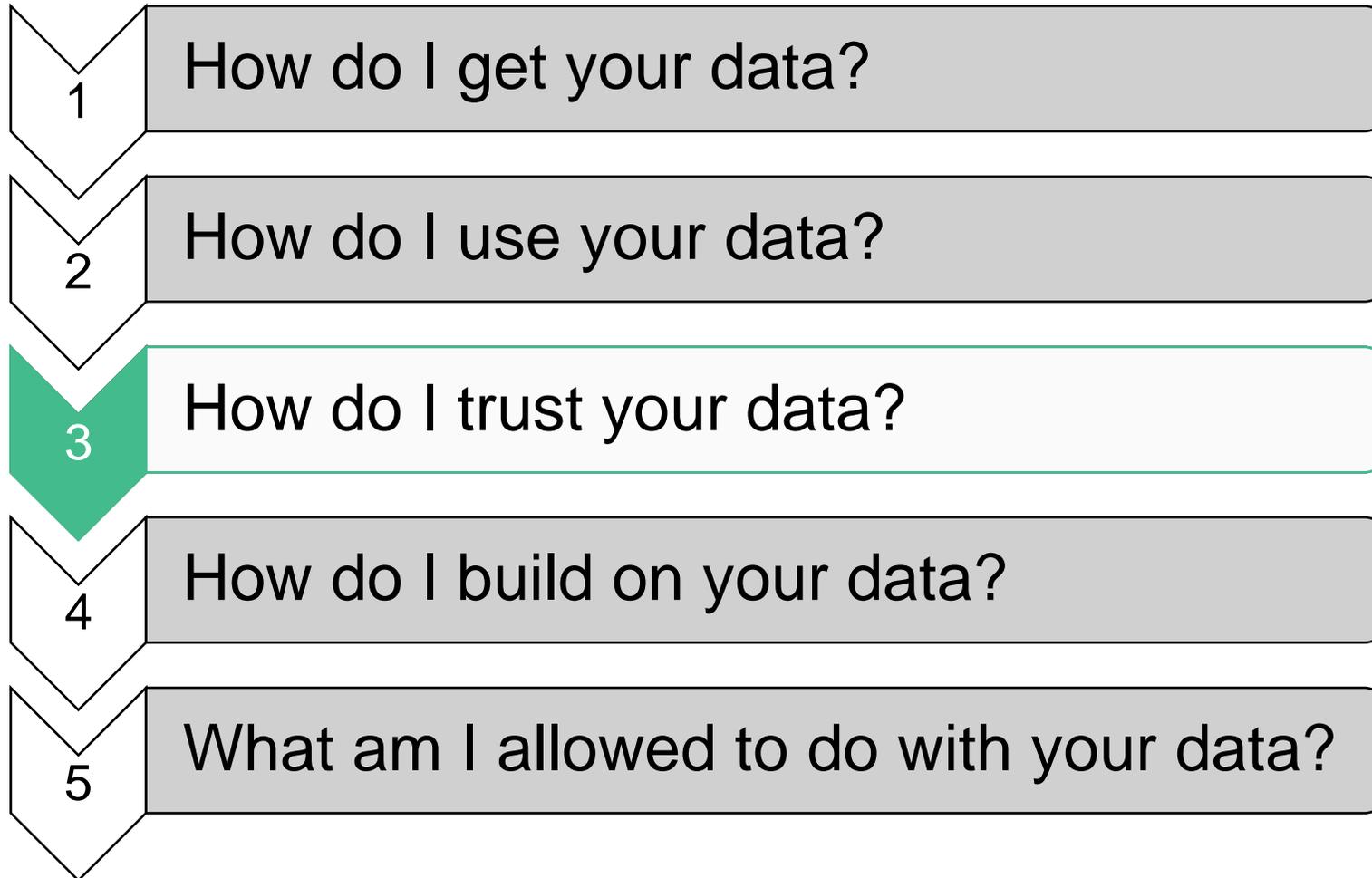
# Filenames and directories

Always seek advice from e.g. your supervisor regarding naming files and directories.

Your group might already have a standard format in place or your supervisor might know what is common in a certain research field.



# What will be discussed?



- Data processing
- Version control

# Data processing

A little bit of detective work....



# Data processing

To trust data, you need to know what happened with it, like a chain of custody.

What was the research question?

Who collected the data?

How was the data processed?

Who owns the data?

All this information is needed before I, or you, can take the data and work with it.

# Data ownership

You mean the data I worked on might not actually be my data?



# Data ownership

Yes, that is exactly what I would like to discuss.

Well, raw data is not owned, but the form in which they are processed is protected; texts, analyses, graphs.



# Data ownership

Who could be responsible for this processed data?

Depending on the agreement with the client the responsibility of the processed data could be with the client (in your case for instance the university) or with the researcher.



# Data processing

- Data storage
  - Data sharing
  - Personal data
  - Documentating data
  - Organizing data
  - Ownership
- Anyonymisation and pseudonymisation
  - Version control



Already discussed



Coming up next

# Processing personal data

Anonymisation, pseudonymisation and personal data

When you work with personal data you have to consider

## **Anonymise your data**

Anonymisation of personal data prevents that the data can be traced back to a person, not even combined with other (anonymous) data.

Try to anonymise the data as quickly as possible during the research project.



See also [data storage](#)

# Processing personal data

Anonymisation, pseudonymisation and personal data

When you work with personal data you have to consider

## **Anonymise your data**

- Remove all direct identifiers  
name, date of birth, (email) address, BSN, telephone number, dates, etc
- Limit/reduce all indirect identifiers  
gender, year/place of birth, body measures, occupation, geographic info, etc
- Remove unusual characteristics or findings
- Reduce level of detail of indirect identifiers
- Delete data not needed

See also [data storage](#)

# Processing personal data

Anonymisation, pseudonymisation and personal data

When you work with personal data you have to consider

## **Pseudonymise your data**

Processing of personal data that the data no longer can be attributed to a specific subject without the use of additional information. For instance a key-file.

See also [data storage](#)

# Processing personal data

Anonymisation, pseudonymisation and personal data

When you work with personal data you have to consider

## **Pseudonymise your data**

Split the data

- Identifiable data (key file)
- Research data required for analysis
- Delete data not needed



Store the key separately from the research data. Limit access to the key, but someone should always have access to this information.

See also [data storage](#)

# Processing personal data

## An example

PERSONAL DATA		ANONYMOUS DATA
<i>Fully identifiable data</i>	<i>Pseudonymous data</i>	
Any data from which a person could immediately be identified with little to no effort (e.g. names, faces, unique characteristics).	Data from which a person cannot be immediately identified, however it remains possible to identify a person from the data.	Data which cannot be traced back to an individual person. Data collected from human subjects, either directly or when re-used for secondary research, is <b>rarely ever</b> anonymous.
<b>EXAMPLE</b>		
Patient number 90210 City Leeuwarden Date of birth 27-4-1967 Income 7.861 Job Judge Car DeLorean License Plate SN-09-HN	Study subject 47110009 Region Friesland Year of birth 1967 Income 7.500-10.000 Job Legal Car DeLorean	Country Netherlands Age 51-60 Income 5.000 – 15.000 Job Legal Car sportwagen
<b>WANT TO KNOW MORE?</b> <b>TURN CARD OVER!</b>	 National Coordination Point Research Data Management <a href="https://doi.org/10.5281/ZENODO.3584842">DOI 10.5281/ZENODO.3584842</a>	 

## Version control

Have you ever analyzed your data and later decided to change your method or code you use for analysis?

# Version control

You are not the only one!

How to keep track of which version you are working with and what analysis was done with which version?



# Version control

In your analysis method you can keep track of changes you make and save each change with a different/new file name → documentation

Small changes are v1.0 to v1.1....v1.x and major changes are v1.0 to v2.0.....vx.0

## Version control

In your analysis method you can keep track of changes you make and save each change with a different/new file name → documentation

Small changes are v1.0 to v1.1....v1.x and major changes are v1.0 to v2.0.....vx.0

My recommendation; document for each file what version of your method you used to analyze the data. Keep track in your document (if possible), filename and a readme file in your directory. This prevents you from having to redo all your analysis because you lost track.

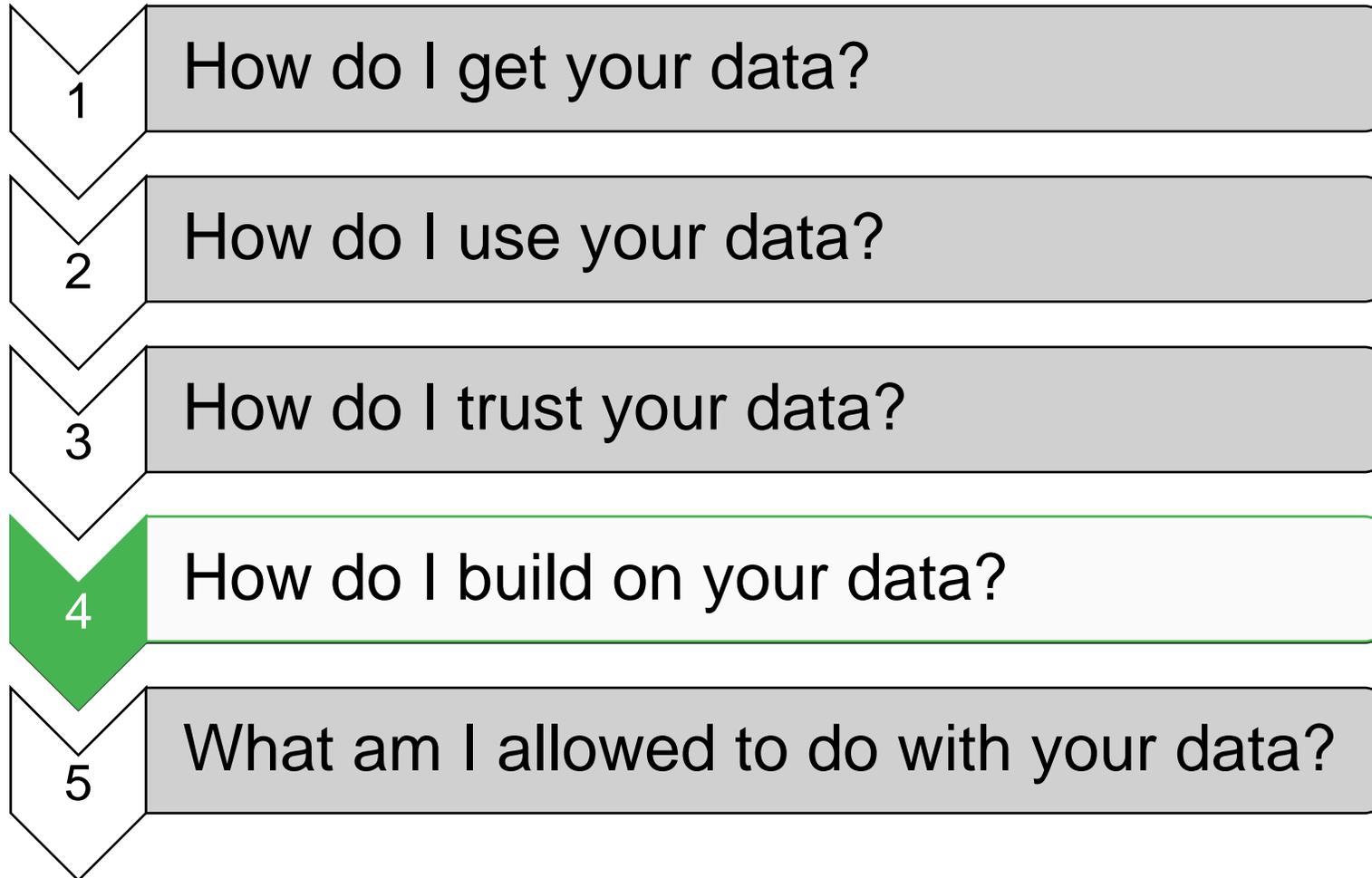
# Version control

In your analysis method you can keep track of changes you make and save each change with a different/new file name → documentation

Small changes are v1.0 to v1.1....v1.x and major changes are v1.0 to v2.0.....vx.0

There is also automatic version control software that could help you stay in control. You can always ask me more about this!

# What will be discussed?



- Data repository
- Data archiving

# Working with someone else's data

If everything is well documented, accessible and compatible, this is simple....

You are pretty much done and ready to go!



## Working with someone else's data

Some interesting research might have been done outside your research group....can you still work with this data?



When a project is completed, the data can be archived in a local archive or a repository such as [DANS](#) and [4TU.ResearchData](#).

Here data sets are stored and, when possible, publicly available for others. You can find datasets and build on these data. Please, check the license associated with the dataset indicating what you can and can't do.

# Archiving your own data

Wait a second! Can I archive my data somewhere when I'm done with a project and where would that be?

When you have completed your bachelor or master project, you can (and should) archive the data .....

*Work in progress*



A short chapter is nice for once, don't you agree?



# What will be discussed?

- 1 How do I get your data?
- 2 How do I use your data?
- 3 How do I trust your data?
- 4 How do I build on your data?
- 5 What am I allowed to do with your data?

- Licenses
- Ethics

# Licenses

When you own a movie, some music or book, you can't share it freely, can you?

Same is true for research data, whether it is the processed data or code used during the project.

You can determine what you can do with the data by looking at the license. In case of your own data, you can require a license, so others have clarity on what they can do with your data.

# Licenses

One important set of licenses are the [creative commons](#)

Let's discuss a few of their options:



No rights reserved; others may freely build upon without restrictions (waive all rights)



You can reuse, adapt and build on the data as long as you give credit to the creator



You can reuse, adapt and build on the data as long as you give credit to the creator and share your data under identical term (with the same license)



You can reuse, adapt and build on the data as long as you give credit to the creator and use the data for non-commercial purposes only



You can copy and share the data in unadopted format only and you give credit to the creator

See also [licensing information from 4TU.ResearchData](#)

# Ethics

One last topic, ethics.



Check whether you need approval of an ethical committee for your project.



When you work with participants, make sure you receive informed consent.

<https://www.utwente.nl/nl/cyber-safety/privacy/leidraad-voor-onderzoek/>

<https://www.utwente.nl/en/organisation/about/integrity/scientific-integrity/ethics-assessment/#review-of-non-medical-scientific-research>

<https://www.utwente.nl/en/organisation/about/integrity/scientific-integrity/ethics-assessment/>

# In summary



When starting a project, take the following items in consideration:

- ✓ Nature of the research data
- ✓ File types
- ✓ Storage during and after research
- ✓ Documentation
- ✓ Ethical aspects
- ✓ Legal aspects (such as licenses)

All done!



When you have questions or would like to have some more information, please contact your [data-steward](#)

Thanks, [Judith](#)

# Thank you

## Images:

- NRC
- HiClipart.com
- Pixabay.com
  - Peggy und Marco Lachmann-Anke
  - Łukasz Siwy
  - Free-Photos
  - Clinker-Free-Vector-Images
  - PublicDomainPictures
  - 200 Degrees
  - 472301
  - dianakuehn30010

## Information:

- Martin O'Reilly, The Alan Turing Institute
- University of Pittsburgh
- Radboud University
- Dans
- ResearchData.4TU
- Landelijk Coördinatiepunt Research Data Management