

cessda eric

Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Data Management Expert Guide

Train the Trainers Workshop
12-13 April 2018, Ljubljana



Sebastian Netscher

Data Archive for the Social Sciences

GESIS – Leibniz-Institute for the Social Sciences



How to cite:

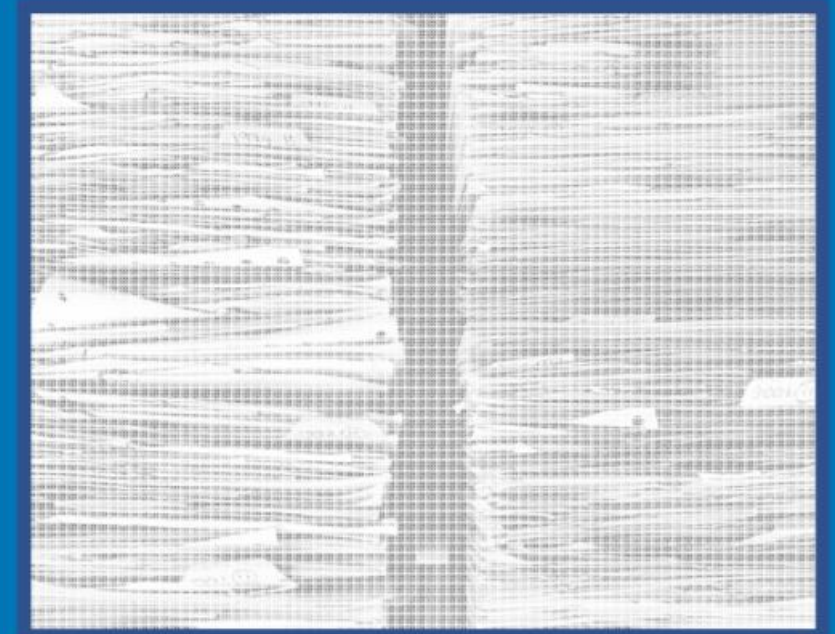
Netscher, S. (2020). CEESDA Data Management Expert Guide. Chapter 4. Store [presentation]. Bergen: CEESDA ERIC.

Storage: What does it mean?

- » Storage: “the action or method of storing something for future use”
- » Data storage: “the retention of retrievable data on a computer or other electronic system”

Oxford Dictionary

- ⇒ Storage does not mean to simply push the storage button, i.e. to put something, somehow, somewhere for future use
- ⇒ Storage is a systematic task



cessda eric

Storage: Why is it of relevance?



» Storage

- various storage solutions
- storage strategy:
 - what is stored and how
 - backup and disaster recovery
 - protect against unauthorized (mis-)use
- part of a systematic data management plan

⇒ Closely connected to other RDM activities, e.g.

- organization (and documentation)
- data protection
- publication (and long-term preservation)

⇒ Requires systematic planning (early task)

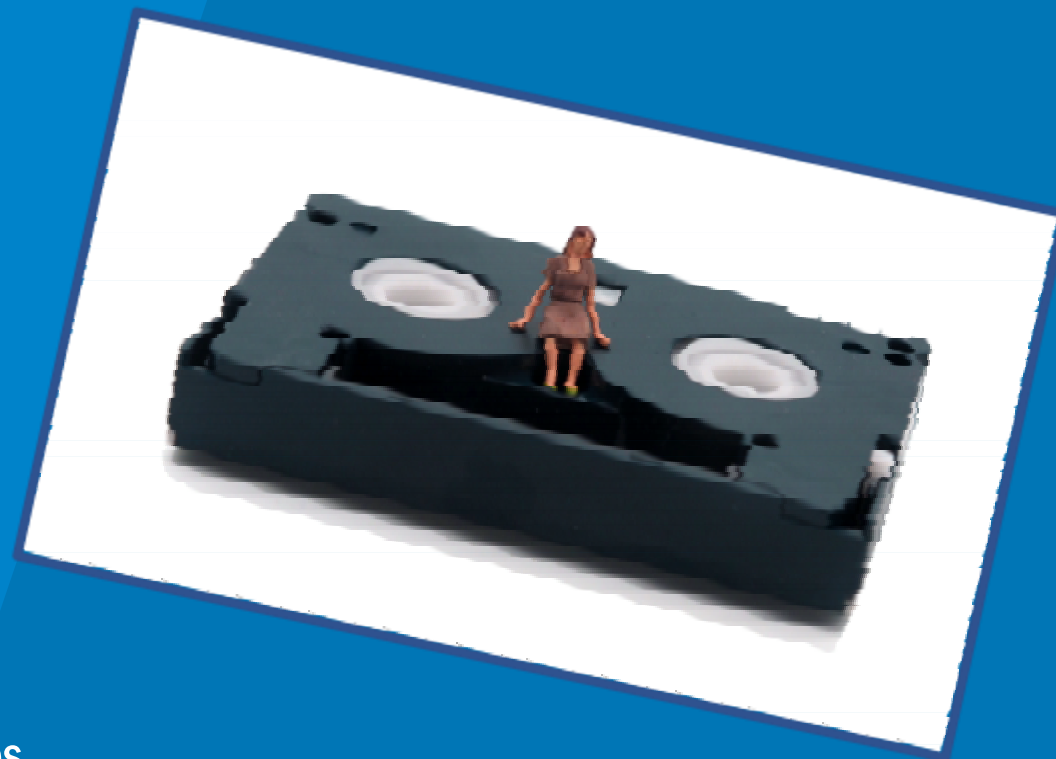
Storage: Some initial questions

- » How much storage space is needed?
- » Who needs access?
- » What precautionary measures should be taken to protect data against loss?
- » Which storage solution suits best?

⇒ No single solution,
depending on particular project, e.g.



- size of data
 - collaborations
 - automatic back-up processes
 - use of sensitive data
- etc.



Towards a Storage Strategy

A Storage strategy contains

- a. storage solutions and media
- b. backup strategy and disaster recovery
- c. data protection

systematically implemented in a data management plan

The Expert Tour Guide

Expert Tour Guide on Data Management

1. Plan
2. Organise & Document
3. Process

4. Store

Storage
Backup
Security
Adapt your DMP: part 4
Sources and further reading

5. Protect

6. Archive & Publish

a. Storage Solutions and Media

cessda eric

Storage solutions: An overview

Portable Devices



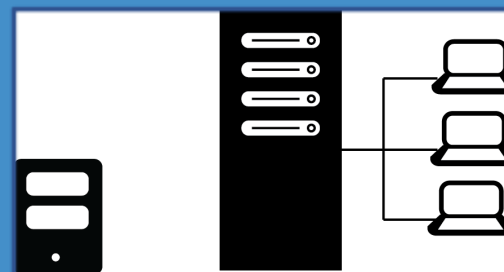
Cloud Storage



Local Storage



Network Drives



Storage solutions: Portable devices

Portable Devices



Advantages

- » easy transport
- » low costs

Disadvantages

- » easy loss, damage etc.
- » not robust for long-term storage
- » problematic quality control

Recommendations

- » encryption and password protection
- » use for temporary storage
- » ensure that device is working

Storage Solutions: Cloud Services

Advantages

- » automatic backups
- » (often) automatic versioning
- » accessible from everywhere

Disadvantages

- » maybe not be suitable for sensitive data
- » insufficient control where data are stored
(2016/679/EC)
- » (partly) loss of IPR

Cloud Storage



Recommendations

- » read terms of services
- » opt for local services
- » encrypt sensitive data

Storage solutions: Local storage

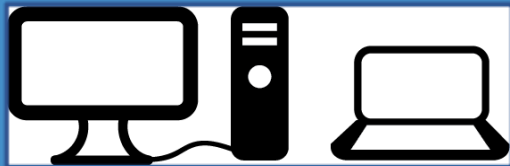
Advantages

- » full control
- » easy to protect sensitive data

Disadvantages

- » high risk of loss
- » accessible only for the one who has access to the computer

Local Storage



Recommendations

- » not suitable for collaborations
- » ensure working with most current version
- » ensure backing up most current version

Storage solutions: Network drives

Advantages

- » central storage
- » shared access
- » central management of backups

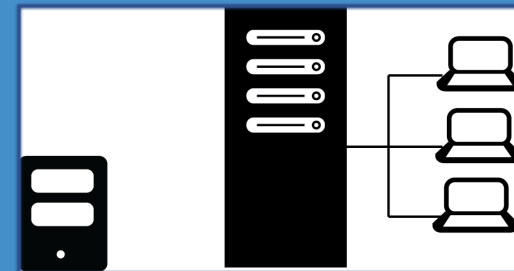
Disadvantages

- » higher security precautions
- » maybe not accessible for external partners
- » higher costs

Recommendations

- » ensure appropriate security strategy
- » ensure versioning control
- » ensure long-term preservation
- » ensure access rights

Network Drives



Storage solutions and media

Optical

- » portable and low costs
- » small capacity, easily damaged and lost, not durable



Portable Flash Drive

- » portable and low costs, robust and long-lived
- » small capacity and easily lost



Some recommendations

- » use at least two types of storage media
- » replace storage media (after 2-5 years)
- » carry out integrity checks, e.g. by checksum tool

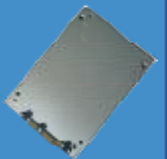
Magnetic

- » low costs and high capacity
- » easily damaged, physical degradation



Build In Flash Drive

- » robust and long-lived
- » high costs and small capacity



b. Backup Strategy

cessda eric

Storage and backups

Various reasons for data loss, e.g.

- » hardware failure
 - » software malfunction
 - » malware or hacking
 - » human error
 - » theft, natural disaster or fire
 - » degradation of storage media
- etc.



* The tweet (Penson, 2017) in the image dates from the 7th of July 2017. Although the tweet is real, the scenario about the contents of the backpack is fictional and based on this scenario in a blog post by Peter Murray-Rust (2011).

Developing a backup strategy

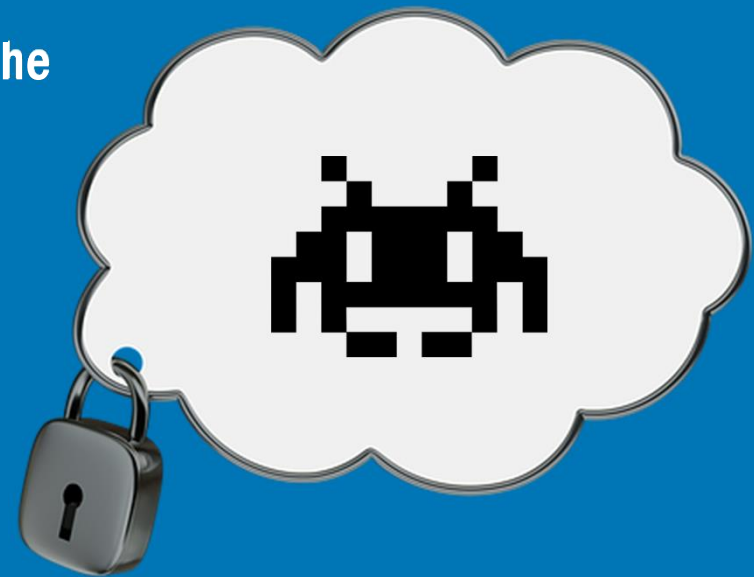
1. Institutional backup strategy
⇒ How does it work?
2. What has to be back upped
⇒ What needs to be copied?
3. When is it back upped and how often
⇒ Frequency and number of copies?
4. Where is it back upped
⇒ Storage solutions for copies?
5. Storage capacity needed for backups
⇒ Memory capacity needed for copies?
6. Tools that can be used to automate backups
⇒ Automate backup processes
of e.g. cloud services?
7. How long is it back upped and how will it be destructed
⇒ Storage period and destruction of irrelevant copies?
8. How will personal data be protected
⇒ Data protection strategy for copies?
9. Disaster recovery plan
⇒ How to access and (re-)use copies
10. Responsibilities
⇒ Who is responsible for backups?

c. Data Protection and Data Security

cessda eric

Storage and security

- » Prevent data from unauthorized access and (mis-)use
- » Ensure not to violate data protection regulations such as the EU's General Data Protection Regulation (2016/679/EC)
- » Data security
 - refers to working files as well as to their backups
 - ⇒ access control, password security, encryption and data destruction
 - ⇒ data security must be closely connected to the overall data protection strategy
 - is supported by various technical measures



Security: Passwords

» A strong password

- has at least 8 to 15 characters
- is a random distribution of characters, combining upper and lower case letters, numerals and special characteristic

» A strong password is never

- used twice
- written down and left lying somewhere
- used in a/n untrustworthy/insecure environment

⇒ Use password generator or pass phrases

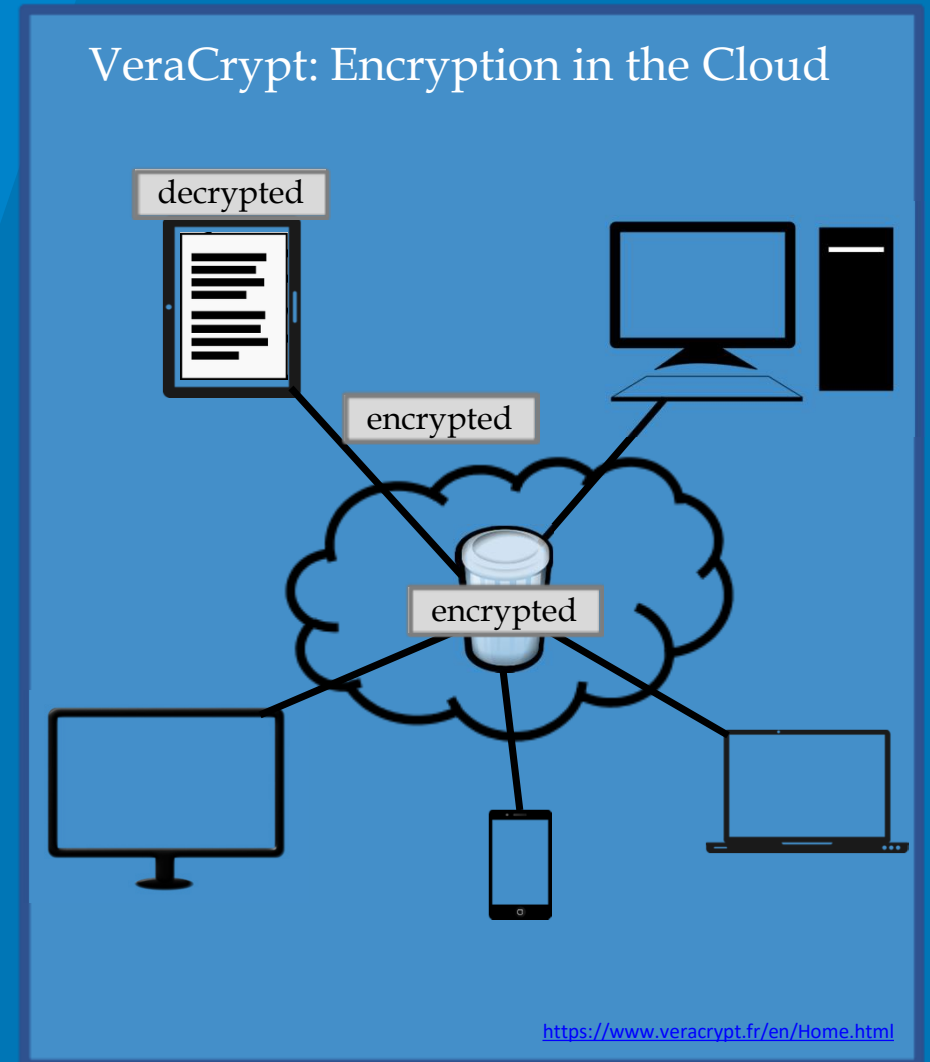
The Worst 10 Passwords of 2017

1. 123456
2. password
3. 12345678
4. qwerty
5. 12345
6. 123456789
7. letmein
8. 1234567
9. football
10. iloveyou

Time Magazine (2018): The Worst 25 Passwords of 2017. Available at:
<http://tim.com/5071176/worst-passwords-2017/>

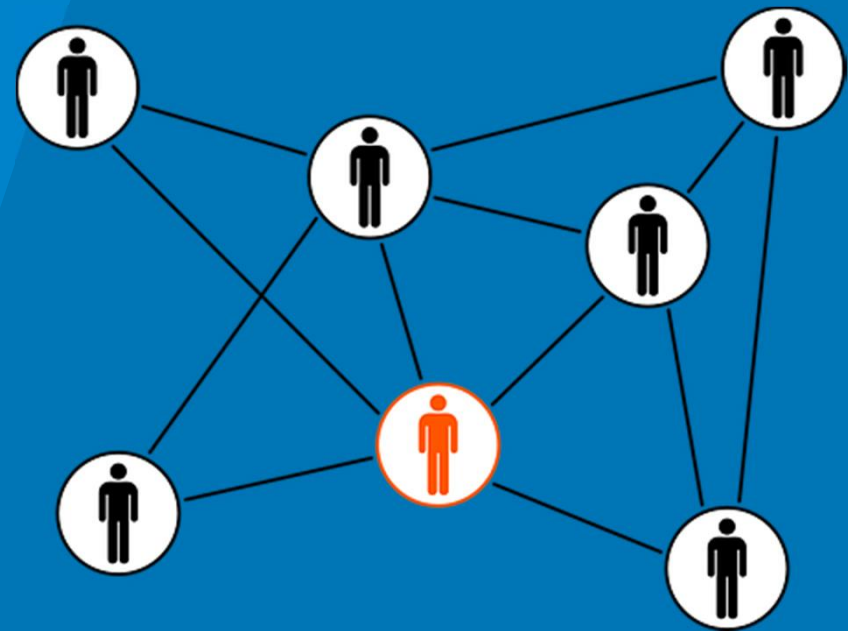
Security: Encryption

- » Encoding digital information in such a way that only authorized parties can view/access it
 - information is “translated” to meaningless codes
 - for back-translation a key (password) is needed
 - ⇒ ensure not to lose the key
 - » Encrypt sensitive information before
 - storing
 - transmitting
 - uploading to the cloud
 - etc.
- ⇒ Various software applications for encryption



Security and research cooperations

- » Define guidelines on data security
- » Restrict data access
- » Encrypt sensitive information before transmitting or uploading to the cloud
- » Ensure not to violate data protection regulation, such as the EU's General Data Protection Regulation (2016/679/EC)



The Storage Strategy and the Data Management Plan

cessda eric

Storage strategy and the DMP

Storage strategy is part of the DMP

⇒ adapt DMP according to

- short-term strategy
 - type of data
 - access regulation
 - storage capacity
 - storage period
 - data security
 - backup procedures
 - budget
- long-term strategy
 - storage period
 - storage location
 - file formats
 - budget



Wrap-up

- » **Storage does not mean to put something, somehow, somewhere**
 - ⇒ requires planning ahead
- » **Systematic storage strategy**
 - ⇒ various storage solutions and media with specific (dis-)advantages
 - ⇒ storage strategy needs to be defined with regard to the particular project:
 - what, how and where things are stored
 - how things are backed up and can be recovered
 - how are things protected against unauthorized (mis-)use
- » **Storage strategy is part of RDM and the DMP**
 - ⇒ adapt DMP according to short- vs. long-term strategy

