How to cite:
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”

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

*Oxford Dictionary*
Storage: Why is it of relevance?

» Storage
  • various storage solutions
  • storage strategy:
    o what is stored and how
    o backup and disaster recovery
    o 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
a. Storage Solutions and Media
Storage solutions: An overview

Portable Devices

Local Storage

Cloud Storage

Network Drives
Storage solutions: Portable devices

**Portables Devices**

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

**Advantages**
- easy transport
- low costs

**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

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

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
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

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

Build In Flash Drive
» robust and long-lived
» high costs and small capacity

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
b. Backup Strategy
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
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

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>123456</td>
</tr>
<tr>
<td>2</td>
<td>password</td>
</tr>
<tr>
<td>3</td>
<td>12345678</td>
</tr>
<tr>
<td>4</td>
<td>qwerty</td>
</tr>
<tr>
<td>5</td>
<td>12345</td>
</tr>
<tr>
<td>6</td>
<td>123456789</td>
</tr>
<tr>
<td>7</td>
<td>letmein</td>
</tr>
<tr>
<td>8</td>
<td>1234567</td>
</tr>
<tr>
<td>9</td>
<td>football</td>
</tr>
<tr>
<td>10</td>
<td>iloveyou</td>
</tr>
</tbody>
</table>

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

VeraCrypt: Encryption in the Cloud

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
Storage strategy and the DMP

Storage strategy is part of the DMP
⇒ adapt DMP according to

• short-term strategy
  o type of data
  o access regulation
  o storage capacity
  o storage period
  o data security
  o backup procedures
  o budget

• long-term strategy
  o storage period
  o storage location
  o file formats
  o 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 define
      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