Chapter 4

Store

Contents

Main take-aways 87
4.1 Storage 88
4.2 Backup 93
4.3 Security 97
4.4 Adapt your DMP: part 4 100
Sources and further reading 102

Main author of this chapter

Jonas Recker, GESIS

CITATION

CESSDA Training Team (2017 - 2019). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. DOI: 10.5281/zenodo.3820473

Retrieved from https://www.cessda.eu/DMGuide

LICENCE

Introduction

The data that you collect, organise, prepare, and analyse to answer your research questions, and the documentation describing it are the lifeblood of your research. Put bluntly: without data, there is no research. It is therefore essential that you take adequate measures to protect your data against accidental loss and against unauthorised manipulation.

Particularly when collecting (sensitive) personal data it is necessary to ensure that these data can only be accessed by those authorized to do so. In this chapter[1], you will learn more about measures to help you address these threats.

Main take-aways

After completing the chapter, you should be:

- Aware of different storage solutions and their advantages and disadvantages;
- Able to plan a storage strategy adequate to the needs of your project;
- Able to plan a backup and disaster recovery strategy to ensure that no data loss, e.g. through human error or hardware failure, will occur during the project;
- Able to decide when and how to protect your data against unauthorised access with strong passwords and encryption.
- Able to answer the DMP questions which are listed at the end of this chapter and adapt them to your own DMP.

1 This chapter is based on information which was put together by the UK Data Service (2017), the online course Research Data MANTRA (EDINA and Data Library, University of Edinburgh, 2017) and Essentials 4 Data Support (RDNL, n.d.).

4.1 Storage

> *I have terabytes of videotaped interviews from a European project, dozens of pseudonymised transcripts and informed consent forms. European partners need access to the files for data analysis. What's the best storage strategy for me?*

A possible storage solution

Type of data	Storage needs	Storage solution
The data which were collected are personal data. Extra security measures to protect it should be in place (see Security).	High storage capacity for videos required; Remote access to videos and transcripts required; Researchers need to work on the same files simultaneously.	Data are transmitted only in encrypted form. (see Security) Data for remote access is stored in cloud storage in Europe. (see Storage) Master copies of videos and transcripts are encrypted and backed up in the cloud and on portable hard disk and flash drives. (see Security) Backups locked away in different, secure locations. (see Backup) Consent forms and encryption keys are stored in a secure safe.

When choosing a suitable storage solution to fit your project's needs, a lot of questions need answering. For example:

- How much storage space do I need?
- Who needs access?
- What precautions should I take to protect my data against loss?
- Which storage solutions are suitable for personal data?

It is an important aspect of data management planning to determine what your storage needs are and select solutions accordingly. In the 'Adapt your DMP' section questions that need answering are covered in more detail.

Storage solutions overview

In the following, you will find an overview of different storage solutions. Factors that play a role are, for example, data sensitivity, ease of access, file size and overall data volume. Advantages and disadvantages are detailed as well as precautions you should take when working with personal (sensitive) data. Each solution closes with recommendations on what to look out for if you decide to use the solution in question.

Portable devices

Advantages	Disadvantages/Risks	Precautions for (sensitive) personal data
» Allow easy transport of data and files without transmitting them over the Internet. This can be especially helpful when working in the field. » Low-cost solution.	» Easily lost, damaged, or stolen and may, therefore, offer an unnecessary security risk. » Not robust for long-term storage or master copies of your data and files. » Possible quality control issues due to version confusion.	Use in combination with encryption and strong password protection.

Recommendations

- Do: use for temporary, short-term storage for non-sensitive data, e.g. in the field or to transport data and files when online transmission is not possible.
- Do: use in combination with encryption and strong password protection, especially if working with sensitive information (see 'Security').
- Do: conduct regular checks to ensure your device is working and that files are accessible.
- Do not: use for long-term storage or master copies of your data and files.

Cloud storage

E.g. Google Drive, OneDrive, Dropbox, a University's OwnCloud, Open Science Framework and Tresorit

Do you want to transfer personal data abroad? Read this first!

The General Data Protection Regulation (GDPR) only permits personal data to be stored within the EU, unless:

- Participants consent to the data being stored in another country (this needs to be real consent i.e. a true choice);
- There are adequate and equivalent levels of data protection in place (e.g. the US/EU Privacy Shield agreement).

However, researchers should assess whether they really need to store the data abroad. If data does need to be stored outside the EU then information sheets and consent forms should clearly identify this and explain the reasons why this is necessitated (See 'Informed consent').

Further guidance on sharing data outside the European Economic Area (EEA) can be found from the Information Commissioners Office.

Advantages	Disadvantages/Risks	Precautions for (sensitive) personal data
» Automatic backups. » Often automatic version control.	» Not all cloud services are secure. May not be suitable for sensitive data containing personal information about EU citizens. » Insufficient control over where the data is stored and how often it is backed up. » Free services by commercial providers (e.g. Google Drive, Dropbox) may claim rights to use content you manage and share them for their own purposes. » Data can be lost if your account is suspended or accidentally deleted, or if the provider goes out of business.	Encrypt all (sensitive) personal data before uploading it to the cloud. This is particularly important to avoid conflict with European data protection regulations if you do not know in which countries servers used for storage and backup are located (see 'Security' for more information on encryption; also see 'Protecting data').

Recommendations

» Do: use cloud services for granting shared, remote and easy access to data and other files to all involved in the project.

» Do: Read the terms of service. Especially focus on rights to use content given to the service provider.

» Do: Opt for European, national, or institutional cloud services which store data in Europe if possible.

  » B2drop (EUdat, n.d.) is an example of a European cloud storage solution.

  » SWITCHdrive (SWITCH, 2017) is a Swiss solution.

  » DataverseNL (Data Archiving and Networked Services, 2017) is an example of a service for Dutch researchers that allows the storage and sharing of data both during and after the research period.

» Do not: make this your only storage and backup solution.

» Do not: use for unencrypted (sensitive) personal data.

Local storage

Advantages	Disadvantages/Risks	Precautions for (sensitive) personal data
» Full control over files. » May be easier to protect against unauthorised access.	» If data and files are stored on only one device, they are vulnerable to loss, e.g. if the device has a malfunction, is stolen or files are overwritten/ erased due to human error. » Only the person who has access to the computer can access the data and files.	Protect the computer with a password and consider encrypting the hard drive.

Recommendations

Using desktop computers and personal laptops as the primary way of storing and accessing data and files is only suitable for projects involving very few people (ideally: only yourself) and where data and files will not have to be moved back and forth between personal computers frequently.

If you plan to work on the data on different (local) workstations, e.g. with your laptop at home and the desktop in the office:

Do: make sure that you always work on the most current version of your files, for example with the help of versioning software or version control guidelines (see 'Data authenticity, versions and editions').

Do: make sure that the most current version is always backed up (see 'Backup').

Networked drives

Advantages	Disadvantages/Risks	Precautions for (sensitive) personal data
» Data and files are centrally stored. » Shared access, remote access for everyone involved in the project possible. » Backups can be centrally managed and automated.	» Higher security precautions are required to prevent unauthorised access and the accidental deletion or manipulation of data and files. » Access for external project partners can be difficult or impossible. » Higher cost.	Use in combination with a suitable security strategy to protect data against unauthorised access.

Recommendations

» Do: Use for distributed collaborative projects involving many people who need access to data and files

» Do: use in combination with a suitable security strategy to protect data and files against unauthorised access (see 'Security').

» Do: use in combination with strict versioning rules (see 'Data authenticity, versions and editions')

» Do: think about long-term archival solutions for data that is complete and has been analysed. Valuable storage space might be released in this way.

» Do: work with rights and permissions to ensure that not everyone has access to everything if this is not required (e.g. access to master files more restricted than access to working files).

Types of storage media

In addition to finding a storage solution that best suits the requirements of your project, you may be required to decide which media types to use for storage and backup of your data and documentation. This is of particular importance if backup and storage are not taken care of by the IT department of your university or research institute.

Optical

Example	Advantages	Disadvantages
» CD, DVD	» Portability » Low cost	» Easily damaged, especially when handled poorly or stored under poor conditions » Easily lost » Frequent read/write errors » Not durable » Relatively small capacity

Magnetic

Example	Advantages	Disadvantages
» Hard Disk Drive (HDD)	» Lower cost compared with built-in Flash drives (Solid State Disks) » High storage capacity	» Subject to physical degradation » Easily damaged (e.g. by magnetic fields or by physical impact)

Flash (portable)

Example	Advantages	Disadvantages
» Solid State Drive (SSD)	» Robustness » Relative longevity	» Data hard to recover if the drive fails » Higher cost compared with magnetic Hard Disk Drives (HDD) » Smaller capacity compared with HDD

Tips for your storage strategy

The UK Data Service (2017b) recommends the following for any storage strategy:

» **Use two types of storage media**
At least two different types of storage media should be used, e.g. Solid State Disk (SSD) and CD-ROM or Hard Disk Drive (HDD) and SDD.

» **Replace storage media**
Replace storage media after 2-5 years.

» **Carry out integrity checks**
Frequently carry out integrity checks to ensure that the stored data has not been corrupted. This can be done with so-called checksum tools. These allow you to detect if a file was changed in any way, intentionally or unintentionally.

How to... check the integrity of your files

We recommend that you frequently check the integrity of your files. This can be done with checksum tools such as MD5summer (n.d.) or Checksum Checker (2014). Such tools create a 'digital fingerprint' - a string of numbers - from the bit values (the ones and the zeros) of a file. Monitoring whether the fingerprint of a given file changes allows you to detect if a file was changed in any way intentionally or unintentionally.

Follow the steps in this video (UK Data Service, 2016b) to perform a checksum check for your own files.

4.2 Backup

Backups are an important instrument to ensure that data and related files can be restored in case of loss or damage. Among the most common causes of data loss are:

- Hardware failure;
- Software malfunction;
- Malware or hacking;
- Human error (research data accidentally gets deleted or overwritten or is lost in transport);
- Theft, natural disaster or fire;
- Degradation of storage media.

Creating a backup strategy in 10 steps

A backup strategy in one sentence would be: Make at least three backup copies of the data on at least two different types of storage media, keep storage devices in separate locations with at least one off-site, regularly check whether they work, ensure you know the process and follow it. In the list below the steps to create a backup strategy are outlined in more detail.

1. Find out whether your institution has a backup strategy

Find out whether your institution has a backup strategy. If so, backups may automatically be taken care of for any files stored on institutional servers. However, it is necessary that you check if the backup strategy in place sufficiently meets your requirements.

2. Determine what you want to back up

The three common options for backups are:

1. Full backup of the entire system and files;
2. Differential backups, where everything is recorded that was changed since the last full backup. To restore your data and/or system, you will require the last full backup and the last differential backup;
3. Incremental backups, where only changes since the last backup are recorded. To restore your data and/or system, the last full backup and the entire series of incremental backups is required.

Differential and incremental backups are also called “intelligent” backups. If only a small percentage of your data changes on a daily basis, it’s a waste of time and disk space to run a full backup every day.

3. Decide how many backups you will need and how frequently to back up

It is recommended that you make three backup copies. This will greatly minimise the risk of data loss, even in the case that one of the backups is damaged or lost. However, if storage capacity is an issue and/or if sensitive data is involved, it may be necessary to work with fewer copies.

You should clearly state in your backup strategy how often backups will be made. The frequency of backups will depend on the frequency and amount of changes to your data and documents.

4. Decide where backups will be stored

We recommend that you store at least some of the backups in (physically) different places. For example, backing up to two servers standing in the same room or building may cause you to lose both backups in case of a fire. Having an offsite copy of your backup mitigates this risk.

Backups can be made to networked drives, cloud storage, and to local or portable devices (see ‘Storage’). What works best for your project depends on the amount of data that needs to be backed up, the required frequency of backups, the level of automation, and the sensitivity of the data.

5. Determine how much storage capacity will be needed

Estimate which amount of data and documentation you will collect and create in your project. Then determine the corresponding approximate amount of storage capacity needed for backups. If your institution has an IT department, they will be able to help you with this.

6. Determine if there are tools you could use to automate backup

Automating backups can help to ensure that backups are created at the correct time and that they are saved to the correct location, reducing the risk of human errors. Both Microsoft and Apple operating systems have software to support automatic backups. Cloud storage solutions too often have a backup functionality. However, make sure to check frequently that functional backups were indeed created.

- OS X
  Have a look at the video tutorial on creating backups for your Mac using Time Machine (UK Data Service, 2016b).
- Windows 10
  Windows 10 includes two different backup programs:
  - File history
    The File History tool automatically saves multiple versions of a given file, so you can "go back in time" and restore a file before it was changed or deleted. That's useful for files that change frequently.
  - Windows Backup and Restore
    The Backup and Restore tool creates a single backup of the latest version of your files on a schedule.

Of course, you would still need an off-site backup as well.

7. Determine how long backups will be kept and how they will be destroyed

It is generally recommended that you do not overwrite one backup with another. However, if you have to back up large amounts of data frequently it may not be feasible to retain all backups for the entire duration of the project.

If sensitive data is involved, make sure that any deleted data are truly gone and cannot be recovered in any way. For suitable procedures, see 'Security'.

8. Determine how personal data will be protected

Make sure that backups of data containing sensitive information are protected against unauthorised access in the same manner as the original files. For suitable measures, see the chapter on Security.

9. Devise a disaster recovery plan

A disaster recovery plan defines the steps to take if a data loss occurred and thus helps you to restore data as quickly as possible. The plan should also assign responsibilities for data recovery tasks and list persons (or functions) to contact when a data loss occurs.

To ensure that data recovery will run as smoothly as possible in the event of an actual data loss, make sure to regularly test whether restoring lost files from your backups is actually possible.

10. Assign responsibilities

Never assume that someone will take care of backups and data recovery. Assign responsibilities for making manual backups, for checking those automatic backups actually happened, for testing data recovery, and for restoring any lost data.

Determine how to check the integrity of backed-up files

Errors can happen when backups are written or copied. We recommend that you frequently check the integrity of your backed up files. This can be done with so-called checksum tools such as MD5summer or Checksum Checker.

The UKDS compares checksums to digital fingerprints. Available tools create such a fingerprint with the help of an algorithm that computes the fingerprint - a string of numbers - from the bit values (the ones and the zeros) of a file. Monitoring whether the fingerprint of a given file changes allows you to detect if a file was changed in any way intentionally or unintentionally.

Video tutorial on using MD5summer: https://www.youtube.com/watch?v=VcBfkB6N7-k

Case studies

In the following, two scenarios will be used to illustrate the importance of backups and to highlight some of the things that are important to consider when planning a backup strategy. After reading through the scenarios, take a few minutes to think about what could have been done to prevent data loss. Afterwards, you can open the tabs to see our diagnosis.

Lost backpack

On a night out after work, a friend's backpack was lost containing literally all of their data and documents for their Master's thesis. A fairly recent copy of the thesis text is backed up in DropBox, but the only two copies of the data - video-recordings and transcripts of interviews with primary school teachers in rural areas of Ireland - were on the laptop (transcripts and sequences from the videos) and the hard drive (original, unanonymised videos and backed-up files from the laptop). Both were lost with the backpack.

Analysis: Measures employed to protect data and participants

- The thesis text was backed up to the external hard drive and to the cloud;
- Transcripts and video sequences from the laptop were backed up to the external hard drive but not to Dropbox because they contained sensitive information;
- No backup of the video footage existed. The entire footage was on the external hard drive in unencrypted form.

Measures that could have reduced the negative effects of data loss

1. Keep backups in different locations

One thing the scenario illustrates is that when it comes to backup, never put all your eggs in the same basket. No matter how many backups you have - if all of them are in the same place, the risk to lose everything is considerable. For storage, consider the advantages and disadvantages of different storage solutions and storage media (see 'Storage').

A rule of thumb is to keep three backups, at least one of them in a different location from the others, on different types of storage media. However, sometimes considerations of privacy or storage capacity will require you to deviate from this recommendation.

2. Use encryption to protect research participants' privacy

In the scenario, the lost hard drive contained personal data of participants in the research. The loss, therefore, compromises the privacy of the involved individuals. Whenever personal data is stored and processed for research, backup measures have to be linked with data protection measures. Personal data should be encrypted and anonymized as quickly and comprehensively as the research objective permits. You should also create only as many copies of this data as absolutely required. Note that this may involve diverging from the "three copies" rule mentioned above.

Master copy gets overwritten

A group of researchers collaboratively works on quantitative survey data. They use a shared working space on a networked drive where a master copy of the data and a working copy are stored. Two backups exist, stored separately from the working and master files: one copy on an external hard drive and one in the university's own Cloud system.

A new researcher enters the project. Who is not aware of the way files are named and organised and accidentally works on the master copy of the data. In this process, a number of variables get overwritten when the new team member recodes variable values and forgets to save them into a new variable. Fortunately, two backups exist.

The researchers know that sometimes copies can get changed due to write or transmission errors, so they decide to check with a checksum tool if the two copies are identical. They discover that the checksums for the two files are not identical. This means that either one or both of the files were altered in some way.

Analysis: Measures employed to protect the data

- The master copy is kept as a separate file from working files;
- Two backup copies on different media and in different locations exist;
- No frequent integrity checks of the backups were made and no additional protection for the master copy of the data was in place to prevent it from being overwritten.

Measures that could have reduced the negative effects of data loss

1. Versioning and file naming rules

Errors such as accidentally overwriting a file can always happen, but they are less likely to occur if clear rules for versioning and file naming are in place and if folders are clearly labeled. Such policies and guidelines help to avoid confusion about what files contain and where they should be saved. See 'Data authenticity, versions and editions'.

2. Restricting access to important files

As mentioned above, human error is one of the most common causes of data loss. Therefore, consider restricting the access to important files, for example with the help of passwords or by using systems with read and access rights management. By giving fewer people access to important files, the risk of data loss caused by human error can be minimised.

3. Creating three backup copies rather than only two

If three copies rather than only two had been created, this would have increased the chances of identifying the unaltered copy: if two out of three copies are identical, this suggests that these are unharmed. This would have saved the project laborious work of trying to identify the correct copy.

4. Checking the integrity of files

Errors can happen when backups are written or copied. These can sometimes make a copy entirely unusable, but sometimes they are small enough to go unnoticed initially but then cause problems further down the line. This could lead to you losing access to the data entirely - for example because a software can suddenly no longer render the files - or it can cause the data to contain errors, thus impacting the results of your research negatively. Learn more about integrity checks in this video about performing a checksum check for your files (UK Data Service, 2016a).

4.3 Security

To prevent unauthorised access and possible changes to your data, data security measures are in order. Such measures, on the one hand, serve to protect personal data and confidential information and on the other hand offer protection against unauthorised manipulation or erasure of files (intentional or unintentional).

Data security can be considerably increased with the help of technical measures. However, these must be accompanied by organisational measures in the form of policies and guidelines.

Measures

In the video linked below, several measures that directly contribute to data security are detailed: limiting access with passwords, encrypting data and disposing of data that you no longer need securely. These measures are exemplified and supplemented by other measures in the boxes below (LSI Storage, 2009).

https://www.youtube.com/watch?v=Ylkg7-JOYX8

Passwords

To protect your data files, you should use passwords to lock the computer systems used to access these data files. The University of Edinburgh (2017) has compiled some guidance on how to choose a strong password. In general, they should be long (15 characters or more). A very useful way to choose strong passwords is to make them up of four randomly chosen and altered words, e.g. C.rr3ctHorseBatteryStaple.

Edward Snowden (LastWeekTonight, 2015) advises us to shift our attention away from passwords to pass phrases which are unlikely to be in a dictionary, e.g. MyMotherM$kesTheB*stCakes. This way of thinking does not only make passwords stronger, but also a lot easier to remember.

The video (Alexanderlehmann, 2015) below explains why pass phrases are hard to crack. It is in German, but you can turn on English subtitles.

https://www.youtube.com/watch?v=jtFc6B5lmIM

Password security

Besides choosing strong passwords, make sure to store and transmit them securely so they cannot be stolen:

- Do: store passwords in a sealed envelope in a secure place (e.g. a safe);
- Do: use secure password management tools. Remembering all of your passwords can be a challenge. Password management tools are one possibility of dealing with this problem. Examples are KeePassX (2017) and Lastpass (2017);
- Do not: write passwords down and leave them lying about openly (e.g. in your desk drawer);
- Do not: enter passwords in untrustworthy environments such as open wifi or internet cafés.

Encryption

Encryption is the process of encoding digital information in such a way that only authorised parties can view it. It is especially useful when you are transmitting personal or confidential data.

When you encrypt a file, the information it contains is "translated" into meaningless code. To translate this code back into meaningful information a key is required. Attacks with ransomware such as the Locky virus ("Locky", 2017) have demonstrated that recovering information from encrypted files without the key is nearly impossible. It is therefore extremely important that you do not lose the key to decrypt your files.

- **Do**: encrypt confidential data, especially before transmitting it online, uploading it to the cloud, or transporting it on portable devices. When working in a team, make sure that the key can be accessed by everyone who needs to access it (but only those people).
- **Do**: ensure that you do not lose the key to decrypt your files, e.g. by keeping it in a sealed envelope in a secure location such as a safe room

Encryption software

The UK Data Service (2017c) has compiled information on encryption and offers short video tutorials demonstrating the use of different software tools to encrypt data.

Commonly used encryption software includes:

- **BitLocker** (2017)
  Standard on selected editions of Windows. For the encryption of disk volumes and USB devices.
- **FileVault2** (Apple Inc, 2017)
  Standard on Apple Macs. For full disc encryption.
- **PGP (Pretty Good Privacy)** (Raicea, 2017)
  There are commercial programmes (e.g. by Symantec (Symantec Corporation, 2017)) and free/open programmes (e.g. Gnu Privacy Guard (GnuPG, 2017)) available.
- **VeraCrypt** (n.d.)
  Multi-platform encryption software (Windows, Mac and Linux). For full disk and container encryption.
- **Axcrypt** (n.d.)
  Open source file-level encryption tool with free and commercial versions available for Windows and MacOS.
- **SafeHouse** (2012)
  Free and commercial software versions available for Windows. Encrypts files, folders and drives.

Physical, network and computer security

To prevent your data from being manipulated or stolen, sufficient security measures to block any unwanted access to rooms and buildings or computers and networks where they are held should be in place.

- Do: log and/or control access to physical sites where sensitive information is stored, e.g. with the help of key cards.
- Do: use strong passwords and encryption (see above).
- Do: use up-to-date virus scanners and firewalls.
- Do: ensure that systems used to access data are continually updated (e.g. security updates for the operating system).

The UK Data Service (2017d) has a list of further important security measures.

Secure disposal

> *Used Phones Are Full of Previous Owners' Data: Researchers bought 20 used smartphones in four cities, and recovered thousands of photos, texts, and emails | Wadell, 2016.*

Managing your data also means thinking about how to securely dispose of confidential information. Merely hitting the "delete" button on your computer or mobile device is not enough. In fact, even formatting the hard drive or doing a factory reset can leave (portions of) confidential information in place.

There are two options for secure disposal of confidential data:

- **The physical destruction of the storage medium** (e.g. shredding of discs)
- **The use of software for secure erasing**
  There are various software options available (UK Data Service, 2017e) that can securely delete files from hard drives. For example, AxCrypt (n.d.), Eraser (2017) and WipeFile (2014) are free open source file and folder shredding utilities.

The UK Data service (2017e) points out that solid-state hard disks (SSD) and USB flash drives (memory sticks) use a different technology than hard drives. Therefore, the techniques for securely erasing files are also different. The use of manufacturer-specific software is recommended. Note, though, that especially for solid state drives and USB flash drives only physical destruction is a 100% guarantee that the data cannot be recovered.

Contact the IT department and the administration of your university or institute to find out about regulations and procedures for secure destruction of confidential data.

Organisational aspects

Data security partly depends on technological and physical protection measures. However, these measures alone are not sufficient and will not adequately protect your data if you do not also address the "human factor". This is particularly important if working collaboratively in a bigger and/or distributed team.

Protection against security breaches depends on the establishment and communication of clear rules and guidelines. Here are some points to consider when planning your data management that focus on the human/organisational dimension of data security:

Do: Invest time to draw up policies and concrete guidelines/checklists for all topics discussed in this chapter, especially:

- Passwords: minimum requirements for password strength; management/secure storage of passwords.
- Encryption: what types of data are encrypted for which purposes using which tools?
- Secure data transmission and transport.
- Secure data disposal.

Do: Restrict access to sensitive data:
Most likely, not everyone on the team needs access to all files. Determine who needs access to which types of data and handle access restrictions, e.g. with the help of passwords. In addition, create a routine to ensure you adapt authorisations in case someone leaves the team.

Do: Create awareness and keep communication going:
Errors often happen due to a lacking awareness of potential issues or threats. For example, does everyone on the team know which data is considered sensitive and why? Is everyone aware of potential risks posed by transmitting unencrypted data via email? Make sure that everyone on the team is adequately involved in discussions of data security issues and measures in place.

4.4 Adapt your DMP: part 4

This is the fourth of seven 'Adapt your DMP' sections in this tour guide. For easy reference, we have put together a list of DMP-questions for all chapters in this tour guide. You can view and download the checklist as pdf (CESSDA, 2018a) or editable form (CESSDA, 2018b), and keep them as a reference while you are studying the contents of this guide.

After working on this chapter, you should be able to define your storage and backup strategy for your data and metadata. To adapt your DMP, consider the following elements and corresponding questions:

Short-term storage strategy

Type of data

Are you collecting personal data or do your data in any other way require special protection?

Who needs access

- Is it necessary to have remote access to the data? Are you e.g. transmitting data from the field?
- How important is fast access?
- Is simultaneous and synchronised access by several people required?

Storage capacity

- How much data are you going to generate and how much storage capacity will you need, including backups?
- Which media types will you use and how often will you replace them?

Storage period

For how long is storage required?

Data security

- How will you protect your data? (passwords, encryption, physical, network and computer security measures?)
- How will the data be disposed of (if need be)?

Backup procedures

- How many backups will you make and where will these be stored?
- How will the integrity of backups and disaster recovery be tested?

Budget

What is the associated cost of storing and backing up data?

Long-term storage strategy

Thinking about storage as part of data management planning also entails considering what will happen to your data and files after the project ends. Where and for how long will data be retained? While the recommended route is to archive and publish your data at the end of the project by handing it over to a data repository, for some data this may not be possible or desired. Maybe no consent was given for sharing, or publication would infringe intellectual property rights of third parties.

To ensure that your data remains accessible even after the end of the project consider the following questions:

Storage period

For how long after the project is the data and the documentation to be kept? 10 years after you last published an article based on the data is commonly considered the minimum period for retention unless legal or ethical issues require shorter or longer retention periods (e.g. see the funder retention requirements of UK funding council policies listed on a Libguide by the University of Southampton (2017)).

Storage location

- Where will the data and documentation be kept after your project ends?
- Can you or your employer guarantee sufficiently secure storage and backup for the data for the envisioned retention period without losing access?

File formats

- Are you certain that your data and files are stored in a format for which there will still be suitable software available to access and process the stored information in ten years?
- Which file formats will you use to minimise the risk that current software can no longer read your data files? See 'File formats' for further information.

Budget

What is the associated cost of storing and backing up data and documentation after the project has ended?

Sources and further reading

Please see the online version of this guide.