Aalto Data Storage Best Practices

This document is a collection of best practices and guidance on how to use the Aalto storage services in a way that supports good data management. We encourage anyone to use this as a template for local adaptations in teams, research groups and departments. These practices originate from the data policy of the Department of Computer Sciences, guidelines of Aalto IT Services and input from Data Agents and other individuals of the Aalto University.

1 Principles for Data Storage

Follow these five principles to store your data for everyone's benefit!

1 Store all research data in shared group directories
2 All data has an owner who makes decisions on access and end of life.
3 Separate working files from data at rest with project and archive folders
4 Aalto provides a reasonable amount of storage to every research group. To get a large amount of storage, > 2 TB, you have to manage your data well.
5 Store your private data only in the **private** folder of your **home** directory. Store any work related files that anyone else could ever need, in another place to ensure their accessibility also in special circumstances.

These principles are common best practices that should fit most of the cases. Exceptions are allowed, but should be considered carefully.

1.1 Use Network Storage Systems

Store all project data in shared group directories.
- All members of the group can access all data.
- Promotes sharing, availability, and long-term continuity.
- Secure storage with automatic backups

Use personal folder or device only for intermediate or temporary files that others won't ever need. Network storage has smaller risk of data loss than personal devices.

Your storage service options are:
1 **Teamwork** group directory, *recommended for research projects*: flexible service supporting large storage needs (> 1 TB per project) and multi-disciplinary projects.
2 **work** group directory (work.org.aalto.fi): suitable for working files that are shared within one department or organizational unit. Has shared quota for whole department.

Contact servicedesk@aalto.fi or your local IT support to get network storage for your group, or to extend your current storage space.

For projects with sensitive data or another reason for more limited access, request another group directory from servicedesk@aalto.fi or your local IT support. Look at the Secure Data Storage Locations guide for more information about working with confidential data.

To collaborate with people outside Aalto University, you'll need to use a file sharing service, cloud storage or collaboration platform. These are described in section 2 below.
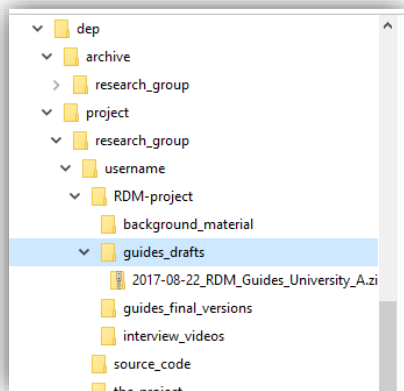
Additional instructions on how to access network storage:
- *Deploying of a network disk in Mac*
- *Deployment of a network drive in Windows*
- *Establishing a remote connection (VPN) to an Aalto network* *(required to access network storage with a personal computer)*

1.2 Organize Data with Folders

There are two top-level folders, `Project` and `Archive`, to separate current working files from past data saved for later use.
- **`Project`** for daily working files, 3-5 years lifetime
- **`Archive`** for longer term storage, 5-10 years lifetime
- Use `Project` while you work actively with your data and after the project, move only the files you need to keep to `Archive`. Plan a date for archival in advance!
- Use `Project` folder for all the data in your research that logically forms a 'project', this may not be the same as a single financial 'project'.

Picture 1: Folder organization, example 1

On the next level, each group has a folder in project and archive. Group members decide how to organize their folder. Consistent structure helps to keep your data organized.

Examples of folder organization

Example 1: Organize first by user, then by subproject:
- `user1/subproj1/…, user1/subproj2/…`
- `user2/…,`
- `user3/…`
- There's a subfolder for each user to help organize the work, but access is shared with all the group members.

Example 2: Organize by project, and use subfolders to separate original and processed data, drafts and final versions, and so on.
- `proj/original`
- `proj/drafts`
- `proj/final`

Organize well from the beginning to keep your data managed! If you don't sort it as you make it, you won't later. Use a separate file to explain the organization strategy for your group / project, so it's easy to keep organized as members of the group change and also easy to find the data afterwards.

1.3 Ownership and Access to Group Directories

Directory owner is a long-term faculty member (professor, supervisor).
- Directory owner makes decisions on data management when needed
- Directory owner makes decisions on access to directory and confirms member addition/removals. All members of the group can access all data. If you need more granularity, IT services will create a new group for you.

Each group must have a succession plan: if owner leaves, someone else has to be able to make decisions on data.
- Defaults to supervisor (of the owner), eventually the head of department
- Most important to define explicitly with groups that have members from different departments.

Directory ownership is for data and access management in Aalto storage systems. Data ownership and intellectual property rights are a separate issue, please check the Aalto IP guide for more information: http://www.aalto.fi/en/about/newsletter_news/2017-09-27-010/

1.4 Owner is Responsible for Clean up

Data must have an expiration time, and this is extended as needed. Directory owner is responsible for the end of life of the data and related management decisions:
- what's kept, is kept for a reason
- what's no longer needed, should be deleted

Owner should clean-up the group folder periodically.
- Use data repositories or CSC PAS (*launch due mid-2018*) for long-term availability of research data

2 Collaborating with Data

There are many services for collaboration. Which one to use, depends on your preferences, collaboration needs and needed security level.

2.1 Cloud Storage Practices

Use Aalto cloud services

Aalto has a service contract for OneDrive for Business, Google Drive and Dropbox. Use these cloud services with Aalto account, and you'll get better terms than with a private consumer account (e.g. more storage space).

Take care of confidentiality

Separate confidential data from the rest. Don't put confidential data to a cloud service without encrypting it first. Look at the Secure Data Storage Locations guide for more information about working with confidential data.

Check the visibility settings for your documents. Google and other search engines may find your word files and presentations in the cloud, if their availability is public.

Comment and review is efficient with the cloud

Instead of using e-mail, upload your file to a cloud service for comments. Share the file with a link, and the recipients may directly comment and edit the file in the cloud.

Data lifecycle and ownership

Owner is needed for data in cloud services as with network directories. Follow the practices in the section 1.3 and make sure to always have a backup person(s) to manage the shared folder.

Before you leave Aalto University permanently, agree with your supervisor where to transfer your work related data stored in the cloud services. Move also any private files to another service.

Aalto University

Data Storage Best Practices

5 (7)

Guide
28.5.2018

V1.0
Public

Learn more about the cloud storage services:
- *[Cloud storage | OneDrive & Google Drive](#) (service descriptions in Inside)*
- *[Cloud services](#) (detailed guides in Inside)*

Here's how to get started:
- *[Deploying of OneDrive for Business](#)*
- *[Google Drive: registration and closing of an account](#)*
- *[Dropbox pilot underway at Aalto University](#)*

2.2 Suggested Collaboration Practices

Here are some general suggestions for collaboration with external partners:

Plan together and document how you work

Use a 'readme' file to document your routines. Include a description of how your workspace is organized and mention any sharing restrictions on files and folders. It's also useful to document software versions used to create files. That will help to access them with a future program version.

When sharing files edited with desktop software (e.g. CAD files), agree on version control practices to prevent conflicting edits: do you work with the same file one at a time or do you keep different files and versions?

Take backups to ensure data availability

Consider taking a scheduled backup of data in collaboration service to another location (for example, synchronize it to your computer or to Aalto network drive). Depending on how much do you trust your collaborators and the service, you might choose to back up everything, only the most important files, or nothing. The more important your data is, more recommended is a local backup.

Share securely

Make sure that everyone knows the sharing restrictions, if you don't want the files to be distributed outside the collaboration environment.

For restricted material, it is safer to require a login instead of granting access to anyone with a link. Link might get forwarded or revealed to third parties.

Cloud storage services work well for sharing files. To collaborate beyond sharing files, have a look at these services:
- *[Eduuni | Teamwork environment](#) (Inside), contact servicedesk@aalto.fi to get a workspace*
- *[Quick guide: Microsoft Teams](#) (conversations, file sharing, etc.)*

2.3 Aalto Network Shares for Departments

Network directory **work** has a folder structure based on departments and organizational units. It's an ideal place for common files of a group, unit or department, where any member of that unit can access them.

The top level folder structure in **work** is based on the cost centre hierarchy maintained by Aalto Financial Services. Each department has code with a letter and three numbers and it's further divided to smaller units that have a code with one letter and five numbers. These folders have access rights that are automatically updated based on the cost centre in the working contract. It's also possible to manually grant access in order to share folders with people from different cost centres and departments.

3 Personal Working Files

Use **home** network directory for the work files that you never share, such as HR files, personal notes and temporary files, etc. Research data or any data for working with team should not be here!

The **home** directory has backups and it's only accessible by yourself. Default quota is 20 GB, if you need more space for individual work, contact servicedesk@aalto.fi to find the best solution for you!

4 Private Data

Each employee is responsible to separate private data properly from work data to protect their privacy. University has the right to manage files located in personal directories in special circumstances, for example to recover work related data when employee isn't available.

Use the folder **private** in your **home** network directory for your private data and non-work related files. To protect your privacy, don't save any private data to other network folders.

Right to manage files is a separate issue from the data ownership and intellectual property rights. Please check the Aalto IP guide for more information: http://www.aalto.fi/en/about/newsletter_news/2017-09-27-010/.

5 Naming Files

Most important is to have a file name that describes the content. Using the same naming convention within the group makes working together easier even further.

Aalto University

Data Storage Best Practices

7 (7)

Guide
28.5.2018

V1.0
Public

Here's one recommended option for naming files:
- `YYYY-MM-DD_filename`
- (alternative: `filename_YYYY-MM-DD`, if you prefer to maintain alphabetic sorting by descriptive name instead of enforcing date sort)

Use these extensions with filename:
- `_vX.X` for basic version management
- `_revXX` after XX has reviewed and accepted the document

6 Enabling Reuse

Organizing, documenting and proper storing increase the value of the data and activities related to it. Following these practices will enable reuse to the original author and the team in Aalto.

Well managed research data will be ready for deposit into a data repository and licensed to reuse by someone else, further increasing the value and the impact of that research.