CSIT Cheatsheet <small>v3.0</small>

Full documentation: **scicomp.aalto.fi** / wiki: **do.cs.aalto.fi**
scicomp.aalto.fi/aalto/welcomeresearchers/

Who?

- CS-IT: Handles CS workstations, CS data, technical procurement, loans of IT supplies. **We are here to help you and experts in facilitating research.**
  - A243 - physical location, guru@cs.aalto.fi - support address, instruction sites at top. wiki.cs.aalto.fi. Chat: scicomp.zulip.cs.aalto.fi/
- Aalto Scientific Computing: Triton (Aalto HPC cluster) and scientific software
  - scicomp.aalto.fi/triton/issues - issue tracker. Chat: see CS above.
  - scicomp@aalto.fi - account requests
  - scicomp.aalto.fi/triton/ - instructions + Scientific computing guides.
- Aalto IT (ITS): Handles Aalto accounts and all basic infrastructure
  - servicedesk@aalto.fi - ITS support address. Non-CS matters.
  - service desks: Main (Otakaari 1M) U133a, W-Th ~08:30-11, 12-14 ; Väre R101, M-T ~8:30-11 and 12-15 (check for updates)
  - it.aalto.fi
  - Aalto Gitlab: version.aalto.fi (If account blocked log in using username@aalto.fi and aalto password. This should unblock HAKA login.)
- CSC: Finland's national computing center, research.csc.fi

Accounts and networks *info: it.aalto.fi*

- Accounts are managed by Aalto ITS, not CS-IT.
- Activation/password reset: password.aalto.fi (w. online bank ID) or srv desks.
- Wifi: aalto open: open network, everyone. eduroam: personal computers, access to protected resources. aalto=for Aalto-managed systems.
- lapa.aalto.fi: CS personal info management.
- Accounts can be extended once you leave, ask supervisor/HR for a visitor contract.

Scientific Computing *info: scicomp.aalto.fi*

- Triton (see above) has 2PB storage, ~10000 CPUs, GPUs, large mem, scientific software. Free, uses Aalto account and Aalto data, scicomp.aalto.fi/triton/
- We have various scientific software on workstations.
- www.aalto.fi/en/services/it-services-for-research - List of common Aalto services.
- Additional computational resources can be found from CSC. Link above.

Data management *info: scicomp → Aalto/Data*, Data/* *

- Aalto provides many free, central, secure, backed-up storage locations. **Avoid local storage when you can!**
- Data should be stored in project directories, not home directories.
- Three main locations: `project` (backed up), `scratch` (large, no backups), `archive`.
- Senior staff can request new projects, see scicomp.aalto.fi. Data can be shared with any number of people.
- Use Aalto Git for version control for code, papers, etc.

Workstations and laptops

- Most are Aalto managed: login with Aalto account, passwords kept in sync. Must log in once on Aalto network before it can work.
- For heavy computation, we usually provide a powerful computer and most work is done on Triton

Linux workstations and laptops *info: scicomp → Aalto/Linux workstations*

- Primary user can install own Ubuntu packages. Use Software Center or:
  - search: `apt-cache search` install: `aptdcon --install PACKAGE-NAME`
- Module system: additional software, `module spider SEARCHTERM` and `module load NAME`. (desktops only). E.g. Python/tensorflow/etc ("`anaconda`"), Matlab, etc.
- Network drives (CS, Triton) available at "`/m/cs/*`".
- NO local data backups: Use network drives instead. Local data lost on reinstall.

Mac laptops *info: scicomp.aalto.fi→ Aalto/Mac*

- Aalto software: Managed Software Center
- Local admin rights for 15 min: Managed Software Center, install and run the "Privileges" app.
- Software installation: run Privileges app (see above) and you can install dmg files
- For advanced software, install brew. See scicomp instructions.

Windows laptops *info: scicomp.aalto.fi → Aalto/Windows*

- Aalto software can be installed from online "Self-service portal". See scicomp.
- To enable admin rights, install the "Enable admin rights for primary user" application from "Start Software installation".. See scicomp instructions.

Remote access *info: scicomp.aalto.fi: Aalto/Remote access*

- `magi.cs.aalto.fi`: CS shell server, all filesystems mounted, can access workstations.
- `{kosh,lyta}.aalto.fi` - all Aalto people, general shell servers. `taltta`: staff
- `vdi.aalto.fi` - Virtual desktop environment, Windows/Linux
- VPN: see Aalto IT instructions. Gets you on Aalto net.
- Aalto nets are: `eduroam`, workstation net, `aalto` wifi *with Aalto laptop* (130.233.*)

Printing *info: wiki: https://wiki.aalto.fi/display/CSdept/Printing*

- Secure Print Aalto account / HSL card. Various locations, check wiki.
- Mobileprint: mail to mobileprint@aalto.fi from Aalto email only, pick up printer.
- Scanning/copying at printers, poster printing available for CS (see wiki)

Common software *info: wiki: IT/Printing (not on scicomp yet)*

- Cloud: Overleaf, Adobe,
- Aalto hosted: Chat (Zulip), Jupyter (jupyter.cs),

Science-IT Data Management Cheatsheet v2.0

Full info: scicomp.aalto.fi/data/

General Aalto info: www.aalto.fi/rdm/

Basics info: *info: scicomp → Aalto/Data Outline*

- There is great value in data, but if it is not handled well, value can be lost. Society now recognizes this and requires good data management all around.
- Funding agencies require good data management for future funding. Start now.
- New projects should request dedicated storage directories.
- **Data should be stored in project directories, not home directories.**
- Projects: Senior staff can request new projects, see docs. To get access to data, find project name, mail and cc data owner for approval. See docs.
- We will accommodate almost any data as long as it is being managed properly.
- Ask us if you have any questions.

Organization strategies

- Separate data by type for proper management. Keeping different types of data separate is the most important step to take at the start.
    - For example: Code vs data, original data vs intermediate files, final results vs other, for-archival vs to-delete, can-be-opened vs confidential.
- Whatever you do, don't copy code. Have master repository in a VCS.
- Back up original data into `/m/$dept/archive/`.
- Traditional project - basic arrangement for one user on one project.
    - `$proj/code/` - code, primary work. Backup to VCS.
    - `$proj/original/` - original data. Backup to archive.
    - `$proj/scratch/` - intermediate files. Replaceable with code+original.
    - `$proj/doc/` - final results, final data, etc. Backup to archive when done.
- Multi-user project - users have dirs organized as above, original data can be shared.
    - `$proj/$user1/…`, `$proj/$user2/…` user's directories as above, code synced with VCS.
    - `$proj/original/`, `$proj/scratch/` - shared files among all users.
- Master project - one project per research group with sub-projects.
    - `$proj/$theme/$user/` - for organized themes.
    - `$proj/$user/$theme/` - user's independent work.

Archival / deleting / opening

- You need to be able to end-of-life your data, otherwise it happens when disks die.
- Large amounts of data can't be stored forever. Separate important from replaceable.
- The easiest way to ensure you always have data is to open it and put in a long-term public repository. Funders encourage this. Zenodo is recommended.

Types of data

Different types of data have different needs. Consider this and keep them separate, so some can be backed up, some deleted. Keep data organized from the start.

- **Code**: backed up and archived, use version control (git checked out wherever).
- **Documentation**: info on how to use code and what you did, similar to code.
- **Original data**: Irreplaceable, at least one copy backed up and archived (project/archive)
- **Intermediate files**: replaceable, closer to computation (scratch, project).
- **Personal data**: Requires special considerations and process, see aalto.fi.
- **Published results**: long-term archived (external repository like Zenode).

Data storage *info: scicomp → Aalto/Data Storage, Data/Data Storage Services*

Not all storage suits everything. Put the right data in the right place.

Qualities: L=large, F=fast, C=confidential, BU=backups, LT=long term backup, S=shareable

Path (on Linux)	L	F	C	BU	LT	S	
$HOME			++				Home dir, 40GB `smb://home.org.aalto.fi/`
/m/$dept/project/$project/	+	+	++	++		++	Typical project files, 10-100s of GB. `smb://tw-cs.org.aalto.fi/project/$project`
/m/$dept/archive/$project/	+	+	++	++	+	++	Per-project archive., 10-100s of GB `smb://tw-cs.org.aalto.fi/archive/$project`
/m/$dept/scratch/$project/	++	++	+	--	--	++	Triton, large, **not backed up**. 10-100sTB* `smb://lgw01.triton.aalto.fi/scratch/$project`
/m/$dept/work/$username/	++	++	+	--	--		Like scratch but per-user.. 10-100sTB* `smb://lgw01.triton.aalto.fi/work/$username`
/l/	+	+++					Local disk storage. Not backed up.
version.aalto.fi			+	++		++	Aalto git repository
Aalto laptops		+	-	--	--		High risk of data loss
External drives, USB, etc			--	---	--		High risk of data loss
$HOME/public_html/				++		++	Webspace (https://users.aalto.fi/~username)
CSC IDA service	++		+		++	++	Long-term archival by CSC
filesender.funet.fi						++	Cloud-based file sending. 50GB
Zenodo (EU project)			--	++	++	++	EU project for long-term open data
Public services (google, dropbox)		-	--		--	+	Be careful, there is no service or confidentiality guarantee.
Google Drive	+	-	--		--	+	Unlimited via Aalto account and registration
Own computers			--	--	--		Unmanaged and risky.

*Scratch and work require Triton accounts.

To mount on own computers via SMB: Ubuntu/Mac "Connect to Server", Win use \\ instead of /. Must use Aalto network instead of VPN.

Version control *info: scicomp: → Aalto/Gitlab, Sci.comp./git*

- At Aalto, private repositories can be hosted an version.aalto.fi
- Code and related data should be in a version control system. Learn one well.
- Git is most common and best supported, but there are others. [git-scm.com](git-scm.com)