

# CSIT Cheatsheet v2.0

Full documentation: <http://scicomp.aalto.fi> / wiki: <http://do.cs.aalto.fi>  
<http://scicomp.aalto.fi/aalto/welcomeresearchers.html>

## 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](mailto:guru@cs.aalto.fi) - support address, instruction sites at top. CS IRC: !cs on IRCnet
- Aalto Science-IT: Triton (Aalto HPC cluster) and scientific software
  - <https://scicomp.aalto.fi/triton/issues> - issues
  - [esupport-triton@aalto.fi](mailto:esupport-triton@aalto.fi) - account requests
  - <http://scicomp.aalto.fi/triton/> - instructions + Scientific computing guides.
- Aalto IT (ITS): Handles Aalto accounts and all basic infrastructure
  - [servicedesk@aalto.fi](mailto:servicedesk@aalto.fi) - ITS support address. Non-CS matters.
  - service desks: Main (Otakaari 1M) U133a, M-F 08:00-15:45, TUAS 1163, M-Th 8:30-11 and 12-15, F 8:30-9:30 and 12-15
  - <https://it.aalto.fi>
- CSC: Finland's national computing center, <https://research.csc.fi>

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

- Accounts are managed by Aalto ITS, not CS-IT.
- Activation/password reset: <https://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.
- <https://basi.cs.aalto.fi> / <https://lapa.aalto.fi>: CS personal info management.

## Scientific Computing *info: scicomp.aalto.fi*

- Triton (see above) has 2PB storage, ~6000 CPUs, GPUs, large mem, scientific software. Free, uses Aalto account and Aalto data, ask for access from Science-IT.
- We have various scientific software on workstations.
- <https://inside.aalto.fi/display/ITServices/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 CS-IT wiki. 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: `apt-get install package-name`
- Module system: additional software, `module spider $searchterm` and `module load $name`. (desktops only). E.g. Python/tensorflow/etc ("anaconda3"), 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/Mac*

- Aalto software: Managed Software Center
- To install any software: Managed Software Center, install "Get temporary admin rights" and you can install programs (dmg files) yourself for 30 minutes.
- For advanced software, install brew. See scicomp instructions.

## Windows laptops *info: scicomp → Aalto/Windows*

- Aalto software can be installed from online "Self-service portal". See scicomp.
- To install most other software, you must apply for a "wa" (workstation admin) account. See scicomp instructions.

## Remote access *info: scicomp: 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. `talitta`: staff
- `rds{01,02}.org.aalto.fi` - Windows terminal server, rdesktop (Aalto nets only)
- VPN: see Aalto IT instructions. Gets you on Aalto net.
- Aalto nets are: eduroam, workstation net, aalto wifi with Aalto laptop (130.233.\*)

## Printing *wiki: IT/Printing (not on scicomp yet)*

- Check wiki for instructions: <https://wiki.aalto.fi/display/CSdept/Printing>
- Secure Print is recommended: print to one queue, pick up anywhere. Identify with Aalto account / HSL card. Various locations, check wiki.
- Mobileprint: mail to [mobileprint@aalto.fi](mailto:mobileprint@aalto.fi) from Aalto email only, pick up at Secure Print printer.
- Scanning/copying: use Secure Print printers.
- Department printers also available: they are cheaper for big jobs.

# Science-IT Data Management Cheatsheet<sub>v2.0</sub>

Full info: <http://scicomp.aalto.fi/data/>

General Aalto info: <http://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 amount 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](http://aalto.fi).
- **Published results:** long-term archived (external repository like Zenodo).

## 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
/m/\$dept/archive/\$project/	+	+	++	++	+	++	Per-project archive., 10-100s of GB smb://tw-cs.org.aalto.fi/archive
/m/\$dept/scratch/\$project/	++	++	+	--	--	++	Triton, large, not backed up. 10-100sTB* smb://lgw01.triton.aalto.fi/scratch
/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.
<a href="https://version.aalto.fi">https://version.aalto.fi</a>				++		++	Aalto git repository
Aalto laptops			--	--	--		High risk of data loss
External drives, USB, etc			--	---	--		High risk of data loss
\$HOME/public_html/				++		++	Webpace ( <a href="https://users.aalto.fi/~username">https://users.aalto.fi/~username</a> )
CSC IDA service	++		+		++	++	Long-term archival by CSC
<a href="https://filesender.funet.fi">https://filesender.funet.fi</a>						++	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 at <https://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. <https://git-scm.com>