



Make the most of the Allas Storage Service

18.4.2023



Welcome to Make the most of the Allas Storage Service webinar!



The webinar will be recorded.

- The Questions session will not be recorded



Please keep your microphone muted.



Please use chat for questions at any time.

- The questions will mainly be answered during the Questions session



During the Questions session you can also raise your hand for asking questions.

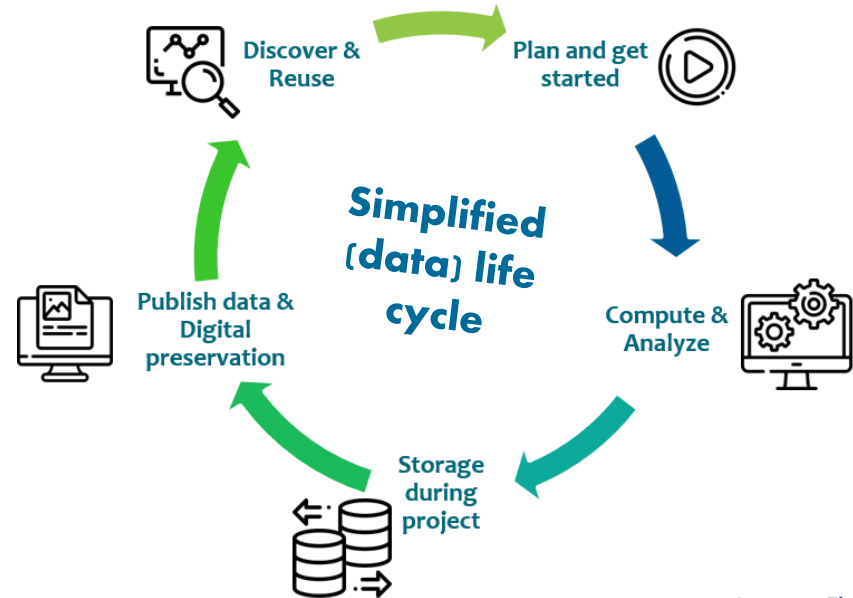
Content



- General introduction - Tua Hindersson-Söderholm
- Demo - Kimmo Mattila
- Time for questions

What do we mean by storing, sharing and publishing data?

- **Storing:** storing data in a service for a period of time
- **Sharing:** sharing data through shared storage space or by using links
- **Publishing** data as dataset: Dataset metadata is available for others via a landing page, that has a persistent identifier
- **Digital preservation:** reliable preservation of digital information for several decades or even centuries



Images: Flaticon.com

Generally about the services' usage policies, the roles and responsibilities

- CSC does not assert ownership or any intellectual property rights to data in our services.
- It's recommended to agree upon the rights to research data within the research group and research organisation early on
- It is the data owner's responsibility to decide which service is suitable for the data in question and that the encryption - if needed - is done appropriately.
- The data owner decides on the openness and usage policies for their data.

Allas Storage Service

- **Purpose of the service:**

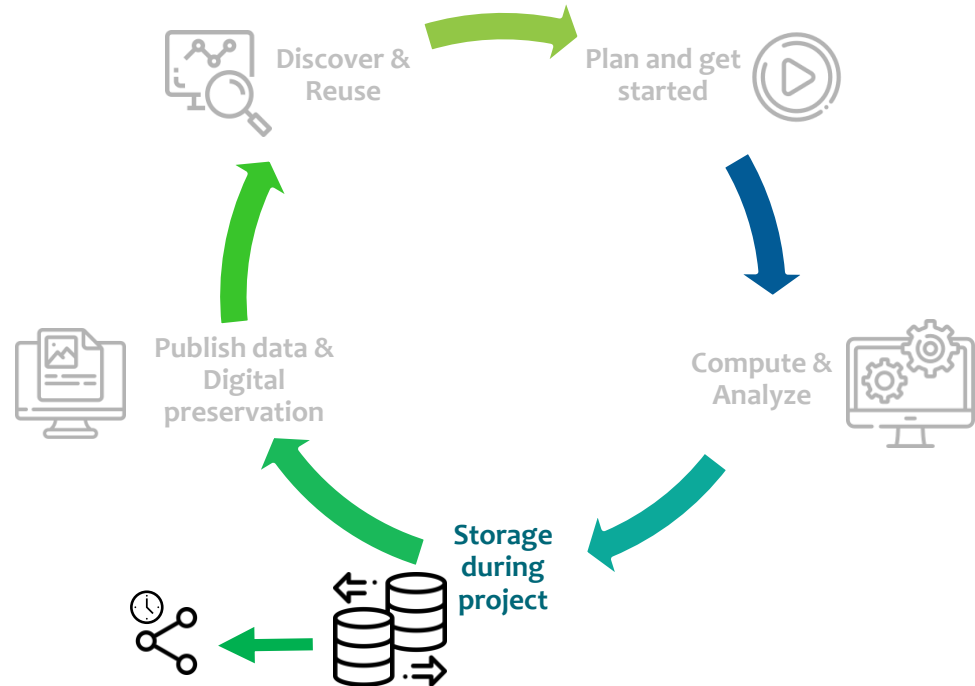
Storing and sharing - especially large amounts of data - during active phase of the research project.

- **Key benefits:**

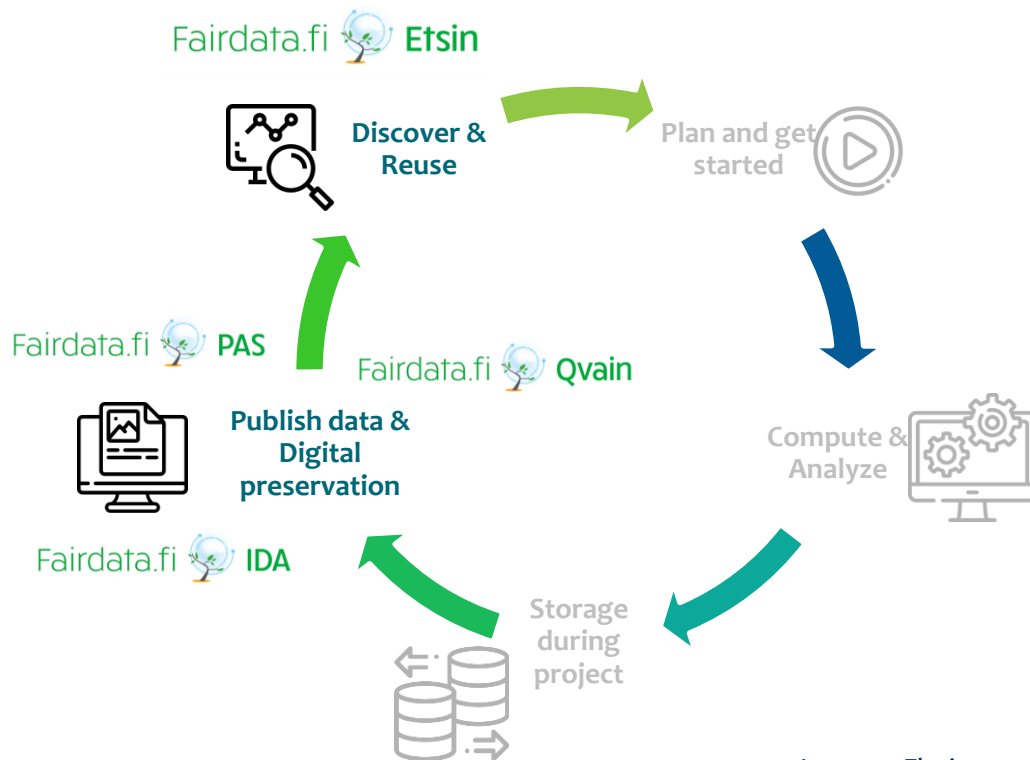
- Data processing can be done using standard APIs from anywhere.
- Data can be shared via public URLs.
- Data can easily be moved to and from CSC's cloud and computing environments.

- **More information:**

<https://research.csc.fi/en/-/allas>

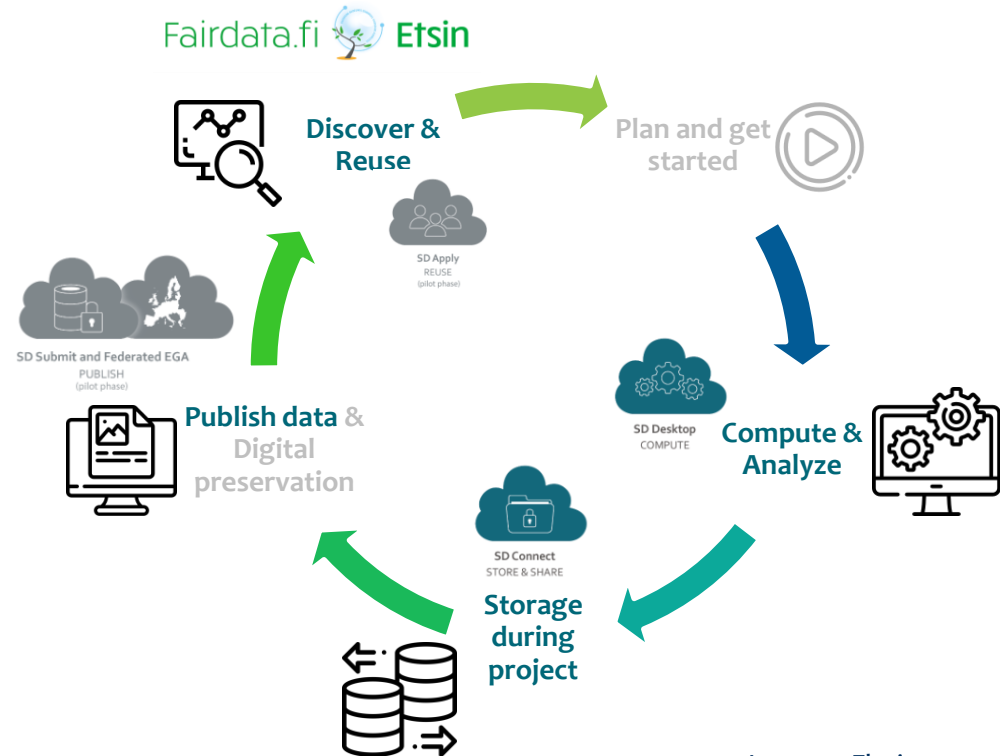


- **Purpose of the service:** Describing, publishing, discovering and preserving research data.
- **Key benefits:**
 - Publishing even large or growing research datasets according to FAIR principles free of charge, with easy to use web tools.
 - Dataset can be transferred to Digital Preservation Service for Research Data.
 - Increased national visibility for published data through [research.fi](https://www.research.fi) portal.
- **More information:** <https://www.fairdata.fi>



Sensitive Data Services for Research

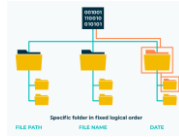
- **Purpose of the service:** Storing, computing, sharing, describing, publishing, and reusing **encrypted sensitive** research data under controlled access (pilot phase).
- **Key benefits:**
 - Available from the public internet, modern web browser (no VPN, no client to install)
 - Data controller has all the tools to manage data access (from data collection to reuse).
 - Data are always encrypted (during data storage, data analysis, data export, and re-use)
 - SD Desktop is a certified computing environment for secondary use of health and social data according to Findata regulations.
- **More information:**
<https://research.csc.fi/en/sensitive-data-services-for-research>



	Allas	Fairdata Services	Sensitive Data Services
Where is the data stored?	In Finland	In Finland	In Finland
Chargeability	Free of charge for Finnish research	Free of charge for Finnish research	Free of charge for Finnish research
Tailoring possibilities	No	No	Limited
Suitable for sensitive data	Yes (encrypted)	<ul style="list-style-type: none"> No In Digital Preservation Service: self assessment to be done by the organization 	Yes (automatic encryption)
Data integrity	Erasure Coding, Checksums	<ul style="list-style-type: none"> RAID, Checksums, File replicas In Digital Preservation Service: part of service promise with e.g. active error correction 	Checksums and/or File replicas in some services

	Allas	Fairdata Services	Sensitive Data Services
Storage period	Active phase of research project	<ul style="list-style-type: none"> Active phase of research project After the project according to home organization policy In Digital Preservation Service: decades or even centuries 	<ul style="list-style-type: none"> Active phase of research project At the end of the research project data can be published for reuse under controlled access
Available storage quotas	<ul style="list-style-type: none"> 10-200 TiB The quota can be increased by a separate application and agreement 	<ul style="list-style-type: none"> 1 Gib to several TiBs In Digital Preservation Service: according to Ministry's decisions 	<ul style="list-style-type: none"> 10-200 TiB The quota can be increased by a separate application and agreement
Sharing data	Yes	Yes, also dataset metadata can be co-edited	Yes (encrypted via SD Connect or via data streaming using SD Desktop)
Publishing datasets with persistent identifiers (PIDs)	No	Yes	Yes
Curation by service provider	No	<ul style="list-style-type: none"> No In Digital Preservation Service: keeping data readable and usable, e.g. conversion of obsolete file formats 	No

File storage



- stores and organizes data as a single piece of information in a folder to help organize it among other data, similar to physical files stored in a paper filing system in an office.
- is also called hierarchical storage.
- when you need access to data, your computer system needs to know the path to find it.

Object storage



- saves files in a flat data environment, or storage pool, as a self-contained object.
- does not actually use directories, folders, and other complex hierarchical organization, even if it can look like that.
- works best for static storage, especially for unstructured data, where you write data once but may need to read it many times.

Allas Essentials

- Create a CSC project in MyCSC and apply for access.
 - Default storage space : 10 TiB
(can be increased up to 200 TiB free-of-charge for the end-user)
 - Add project members – all with the same rights to all data
 - Data can also be shared to non-project members with public URLs
- An object cannot be modified while it is in Allas. It must be downloaded to a server for processing, and the previous version replaced with a new one.
- Limits for number of buckets and objects are fixed by CSC policy.

The default quotas for projects:

Resource	Limit
Storage amount	10 TiB
Buckets per project	1 000
Objects per bucket	500 000

Planning how to organize the objects in Allas pays off in the long run

- Some workflows and software create complex directory structures to store and manage data, resulting in thousands or even millions of individual files.
 - Note that storing a large number of **small** objects will cause performance degrading in Allas.
- In principle, all files could be stored as separate objects to Allas, but in that case, objects must be split into multiple buckets giving you tenths or hundreds of buckets.
 - Note the **limits** of 0,5 million objects/bucket and 1 000 buckets/project.
- As another extreme option, all files could be collected into one archive object.
 - The names of the **individual** files included in the object is added to the object metadata.
- Packing at a higher level in the hierarchy is often the optimal way of storing in Allas.
 - Consider for example to pack the data by device and/or by year etc.



Support: servicedesk@csc.fi



facebook.com/CSCfi



twitter.com/CSCfi



youtube.com/CSCfi



linkedin.com/company/csc---it-center-for-science



github.com/CSCfi