



CSC

ICT Solutions for  
Brilliant Minds



## Sensitive Data at CSC

Francesca Morello, Martin Matthiesen, CSC 23.9.2022



# Contents

- Introduction
- The SD Toolbox
- Key features
- SD Connect: store and share
- SD Desktop: Compute
- SD Publish: Make available
- SD Apply: Request Access
- Summary

# The SD Toolbox

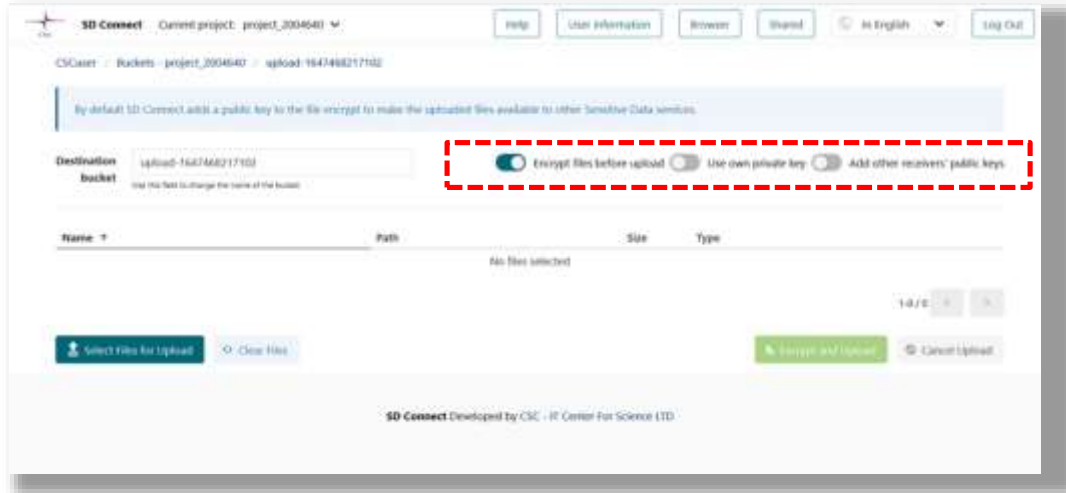
- SD Connect: The encrypted data pipeline and storage
- SD Desktop: A virtual desktop into an isolated VM
- SD Submit: The metadata publisher
- SD Apply: links between
  - Data owner
  - Data user
  - Data access committee
- Development funded by Elixir Finland and NeIC
- Operation funded by Finnish Ministry for Education & Culture



## Key features of the Service

- **(Relatively) easy to use:** web user interfaces
- **Easy to access:** available from the public internet using MFA, modern web browser (no VPN, no client to install)
- **Available on demand:** my.csc.fi portal
- **Data controller** has all the tools to manage data access
- **Secure data management:** data is always **encrypted** (during data storage, data analysis and re-use)

# SD Connect: store and share



- **Default:** automatic, integrated encryption during upload using the Sensitive Data encryption key
- **Optional:** multiple encryption keys from
  - Researcher
  - Collaborator

## SD Desktop: compute

- SD Desktop: A remote desktop into a pre-built VM not otherwise connected to the internet.
- Requires Multi-Factor Authentication, MFA.
- Available options offer a range of capabilities from doing simple statistical analysis to machine learning.
  - Open Office
  - R-Studio
  - Python
- Customisable via singularity containers.

# SD Desktop: compute



- Isolated from the internet (copy/paste disabled)
- Analyse encrypted data stored in SD Connect (via Data Gateway)
- Data export (CSC project manager): results can be **exported** using a specific application (airlock, only CLI)
- Can be **deleted or launched on demand**
- Findata audited SD Desktop available

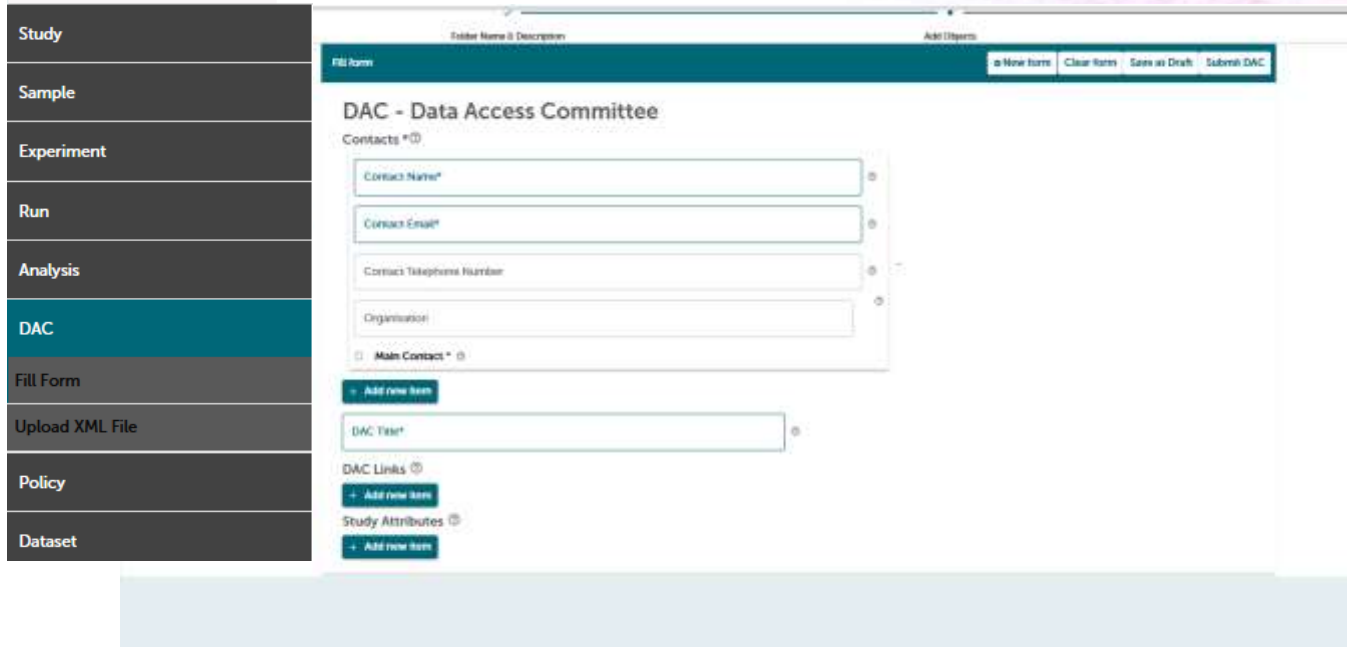
## SD Submit: publish under controlled access

- Dataset is uploaded to CSC via web interface or API
- Dataset is described with appropriate metadata
  - Currently biology is supported best
  - Essential for Discovery & Re-use
- Legal agreements currently not automated
- Each dataset linked to :
  - policies which regulate its re-use, complying with data subject consent
  - Data Access Committee (representing the Data Owner)



# SD Submit: DAC metadata object

- Link data to a the Data Access Committee that will decide who can access them



The screenshot displays the 'SD Submit' web interface. On the left is a vertical navigation menu with the following items: Study, Sample, Experiment, Run, Analysis, DAC (highlighted in teal), Fill Form, Upload XML File, Policy, and Dataset. The main content area is titled 'DAC - Data Access Committee' and contains the following fields and controls:

- Buttons: 'New Items', 'Clear Items', 'Save as Draft', 'Submit DAC'
- Section: 'Contacts \* (0)'
  - Form fields: 'Contact Name\*', 'Contact Email\*', 'Contact Telephone Number', 'Organisation'
  - Section: 'Main Contacts \* (0)' with an 'Add new item' button
- Form field: 'DAC Title\*'
- Section: 'DAC Links (0)' with an 'Add new item' button
- Section: 'Study Attributes (0)' with an 'Add new item' button

# SD Apply: Requesting access

Catalogue Applications Martin Mathiesen Sign out

## Application 2022/35: testing

Fill out and submit the application. Several applicants can be invited as members of the application, and each applicant has to accept the terms of use to get access.

### State

Apply
Approval
Approved

Application: 2022/35

Title: testing

Status: Applied

Last activity: 2022-03-08 13:30

### Events

2022-03-08 13:30 Martin Mathiesen submitted the application for review.

2022-03-08 13:30 Martin Mathiesen applied a draft.

2022-03-08 13:30 Martin Mathiesen accepted the terms of use. [Show more](#)

### Actions

[Copy as a new application](#) [Edit](#)

### Applicants

**Applicant**

Name: Martin Mathiesen

Accepted terms of use:

[Show more](#)

### Resources

**1000 Genome Dataset – More info** [🔗](#)

### Terms of use

Each member is required to accept the terms of use to access the resources.

Creative Commons Attribution 4.0 International License [🔗](#)

Terms of use accepted

### Application

**Base form for registering a resource**

Title of the application (max 250 characters) (optional)

testing

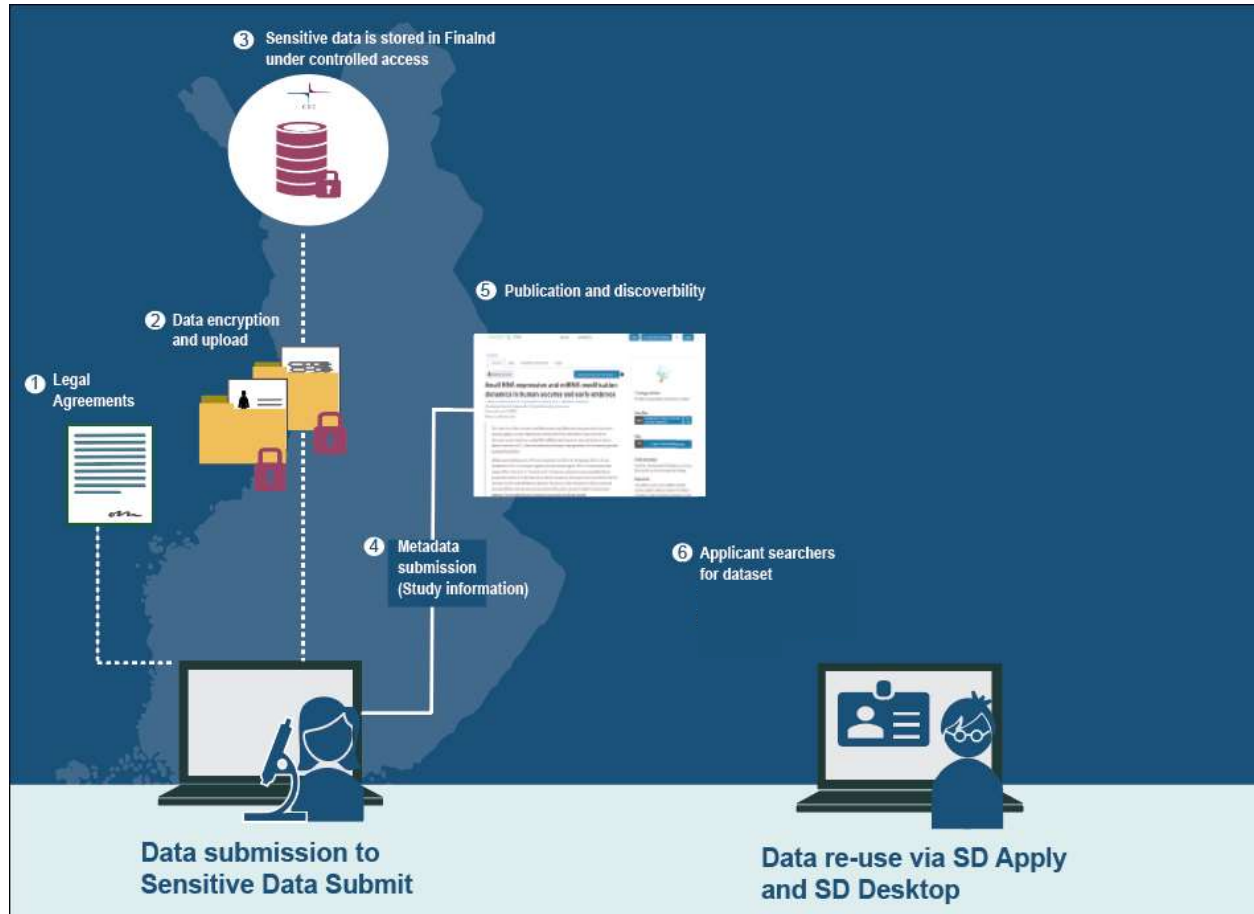
Select type of submission:

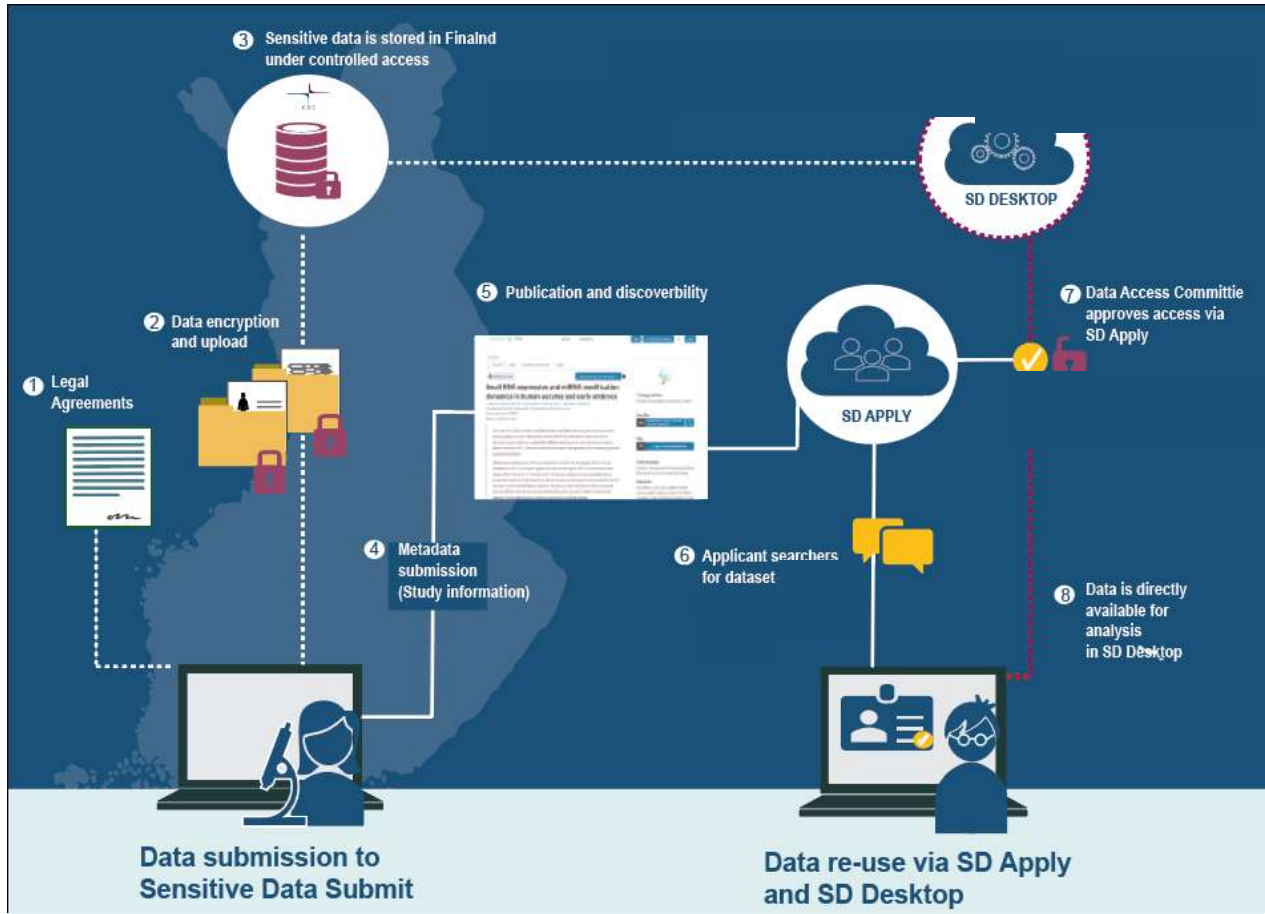
SD4 stand-alone

Stand-alone SD4 archive (see case description) (max 200 characters)

test







## Further work

- Automated flexible key management
  - Users should not need to handle keys
- Using the super computers for secure processing
  - Public beta for the national super computer Puhti
  - Proof of Concept for the upcoming LUMI



**Thank you!**

For more see:  
<https://docs.csc.fi/data/sensitive-data/>



[facebook.com/CSCfi](https://facebook.com/CSCfi)



[twitter.com/CSCfi](https://twitter.com/CSCfi)



[youtube.com/CSCfi](https://youtube.com/CSCfi)



[linkedin.com/company/csc--it-center-for-science](https://linkedin.com/company/csc--it-center-for-science)



[github.com/CSCfi](https://github.com/CSCfi)