



CSC

ICT Solutions for
Brilliant Minds



Sensitive Data at CSC

Francesca Morello, Martin Matthiesen, CSC 11.5.2022

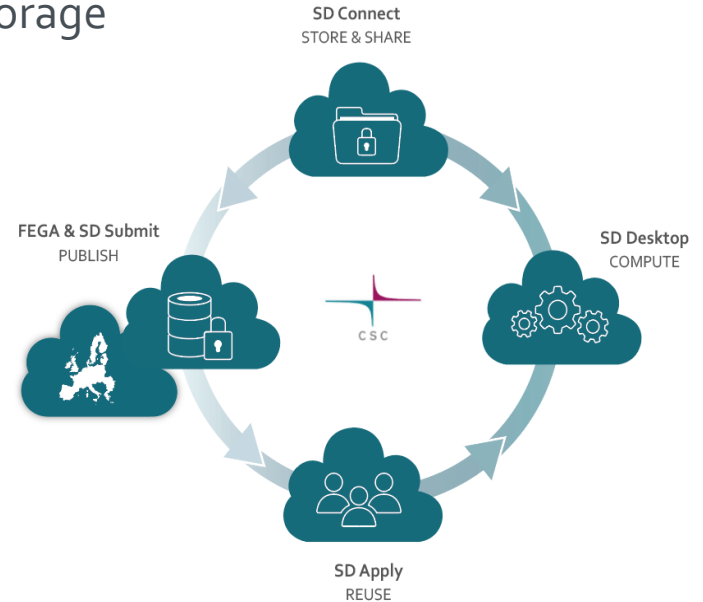


Contents

- 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

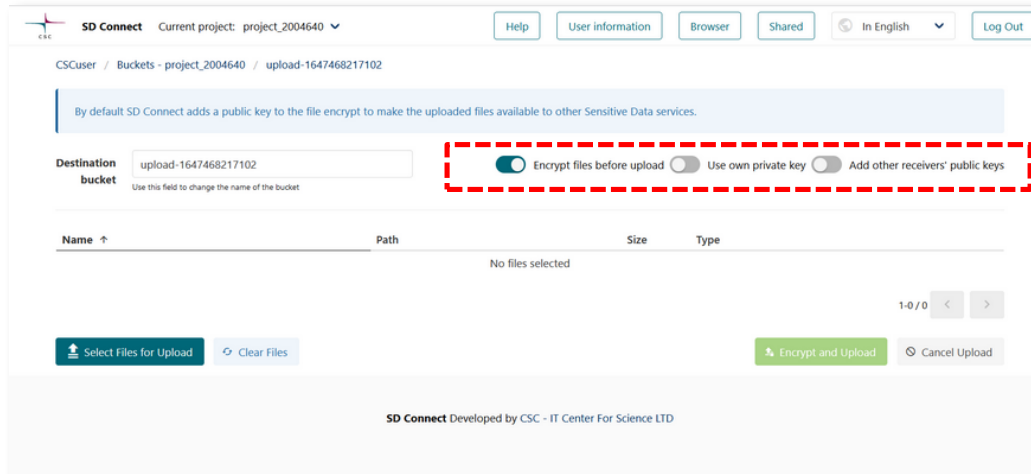
- SD Connect: The encrypted data pipeline and storage
- SD Desktop: A virtual desktop into a closed VM
- SD Submit: The metadata publisher
- SD Apply: links between
 - Data owner
 - Data user
 - Data access committee
- Development funded by Elixir Finland and NeIC
- Operation funded by Finnish Ministry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

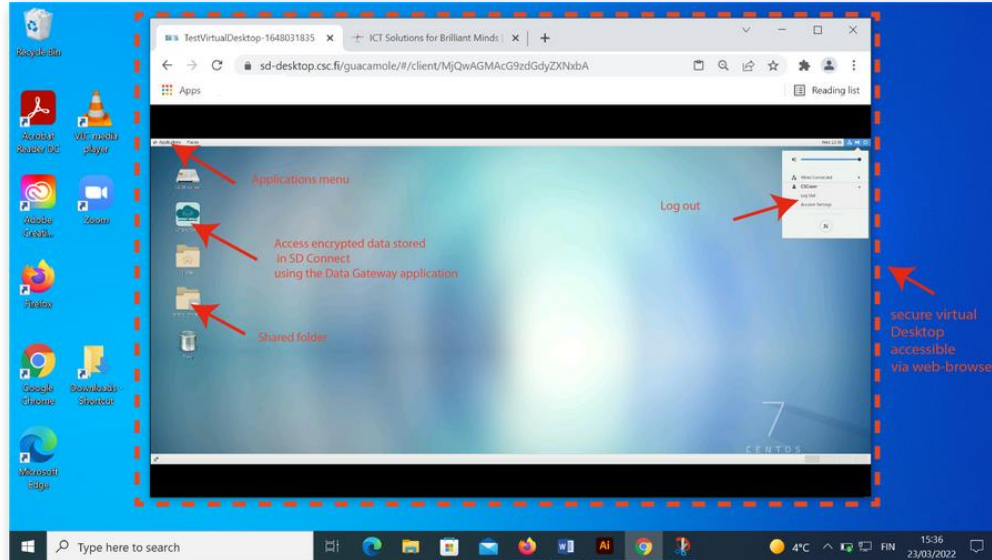


- **Default:** automatic, integrated encryption during upload using the Sensitive Data encryption key
- **Optional:** multiple encryption keys from
 - Researcher
 - Collaborator

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



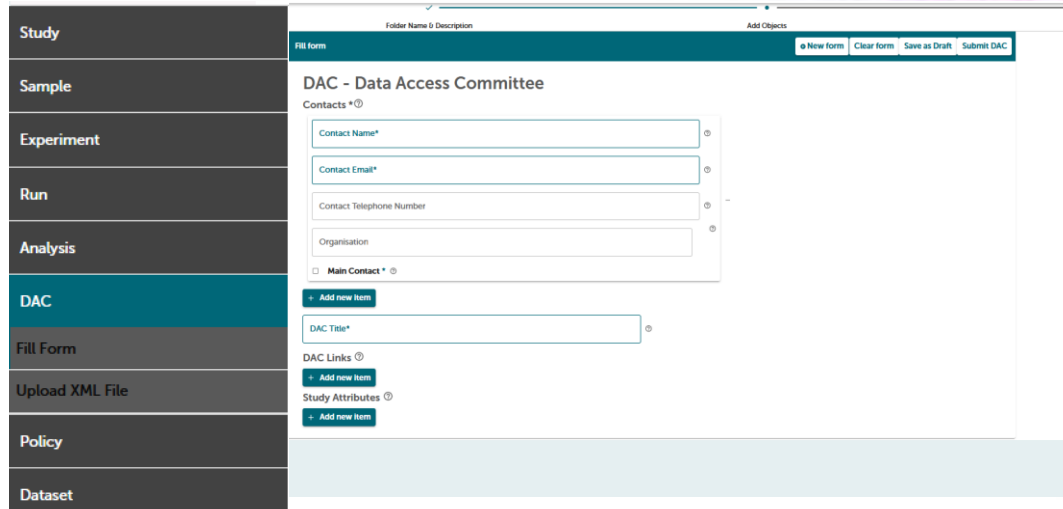
- Isolated from the internet (copy/paste disabled)
- Analyse encrypted data stored in SD Connect (via Data Gateway)
- Data export (CSC project manager): results can be **exported** using a specific application (airlock, only CLI)
- Can be **deleted or launched on demand**

SD Submit: publish under controlled acces

- Dataset is uploaded to CSC programmatically (ssh connection)
 - Automatic datatransfer from SD Connect to SD Submit is WIP
- 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 shows a web interface for submitting DAC metadata. 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 sections:

- Contacts *** (with a plus icon):
 - Contact Name* (text input)
 - Contact Email* (text input)
 - Contact Telephone Number (text input)
 - Organisation (text input)
 - Main Contact * (checkbox)
- +** Add new item (button)
- DAC Title* (text input)
- DAC Links** (with a plus icon):
 - +** Add new item (button)
- Study Attributes** (with a plus icon):
 - +** Add new item (button)

At the top of the form area, there are buttons for "New Form", "Clear Form", "Save as Draft", and "Submit DAC".

SD Apply: Requesting access



Catalogue Applications

Martin Matthesen 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 gain access.

State

✓ Apply → Approval → Approved

Application	2022/35
Title	testing
State	Applied
Latest activity	2022-05-06 15:50

Events

2022-05-06 15:50	Martin Matthesen submitted the application for review.
2022-05-06 15:50	Martin Matthesen saved a draft.
2022-05-06 15:50	Martin Matthesen accepted the terms of use.

[Show more](#)

Applicants

Applicant

Name	Martin Matthesen
Accepted terms of use	<input checked="" type="checkbox"/>

[Show more](#)

Resources

Actions

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

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 (CC BY 4.0)

Terms of use accepted.

Application

Base form for registering a resource

Title of the application (max 100 characters) (optional)

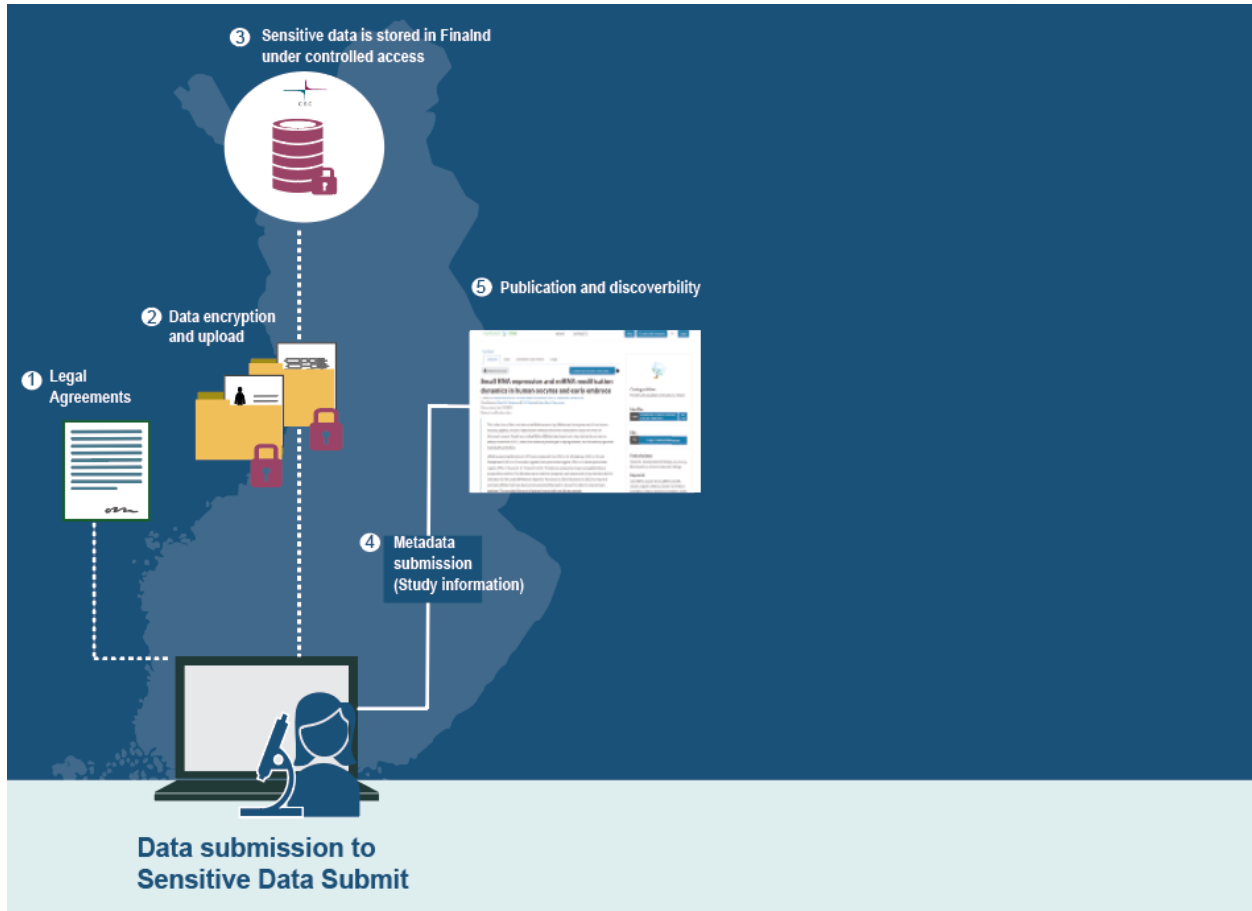
testing

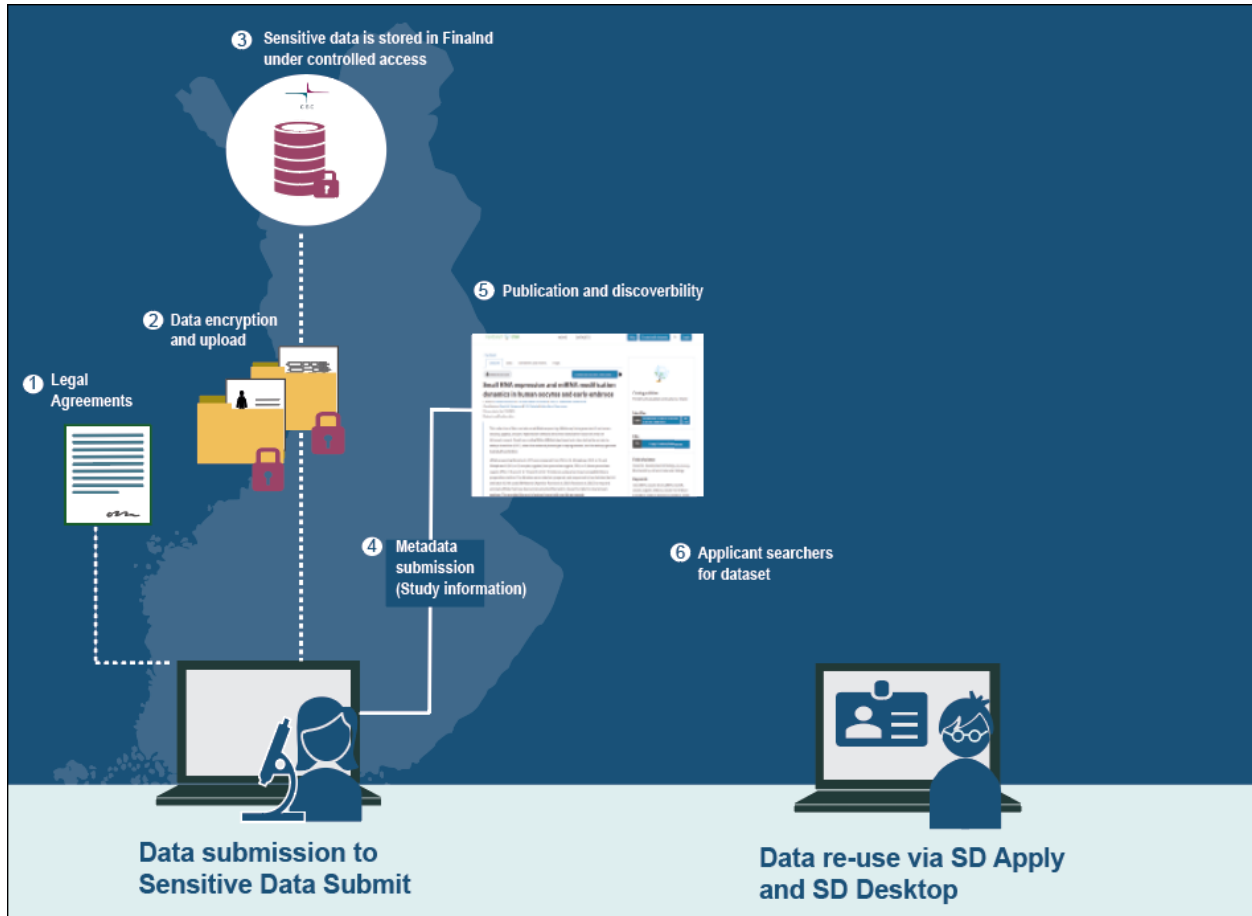
Select type of submission

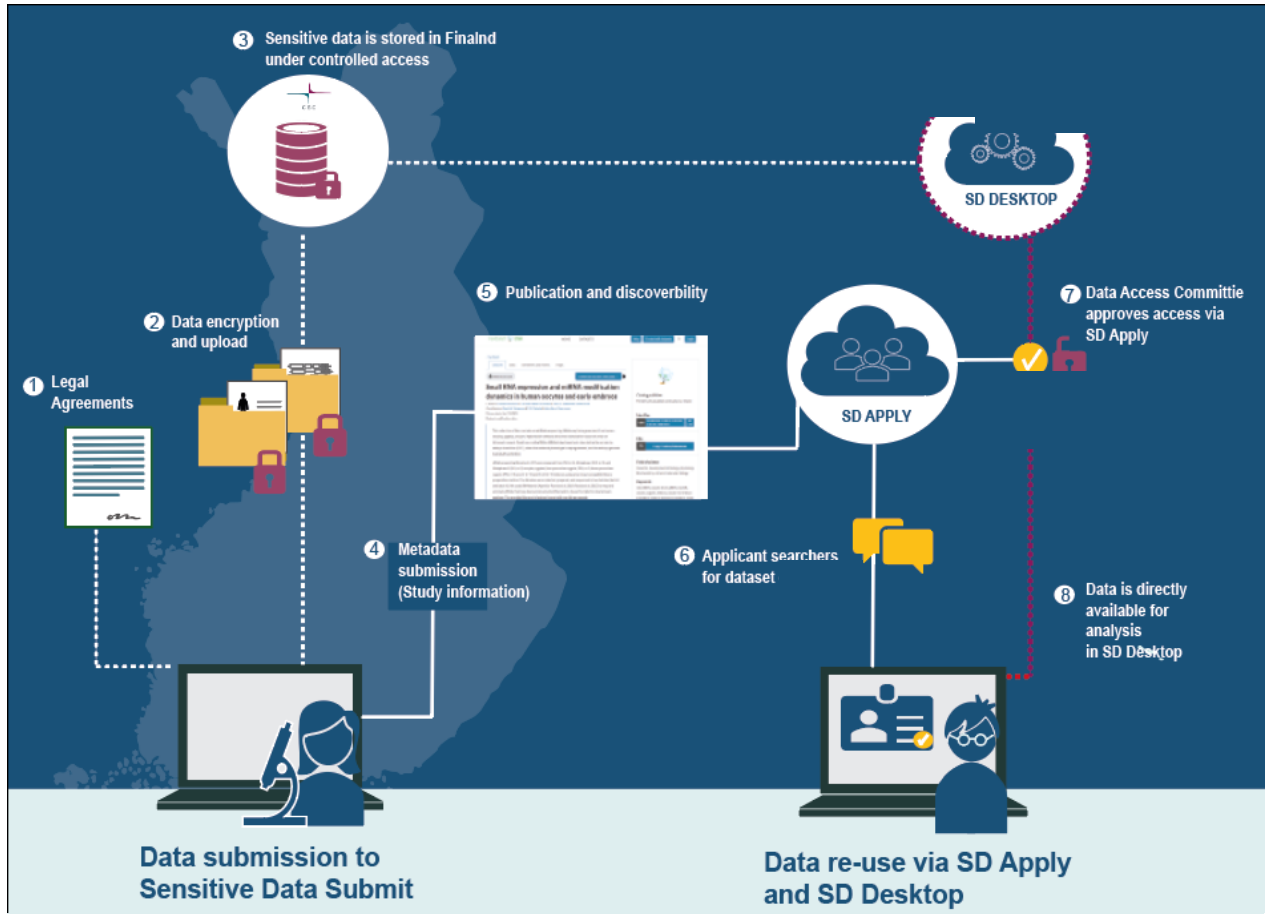
SDA stand-alone

Stand-alone SDA archive use case description (max 200 characters)

test









Thank you!

For more see:

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



facebook.com/CSCfi



twitter.com/CSCfi



youtube.com/CSCfi



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



github.com/CSCfi