CSC

ICT Solutions for Brilliant Minds

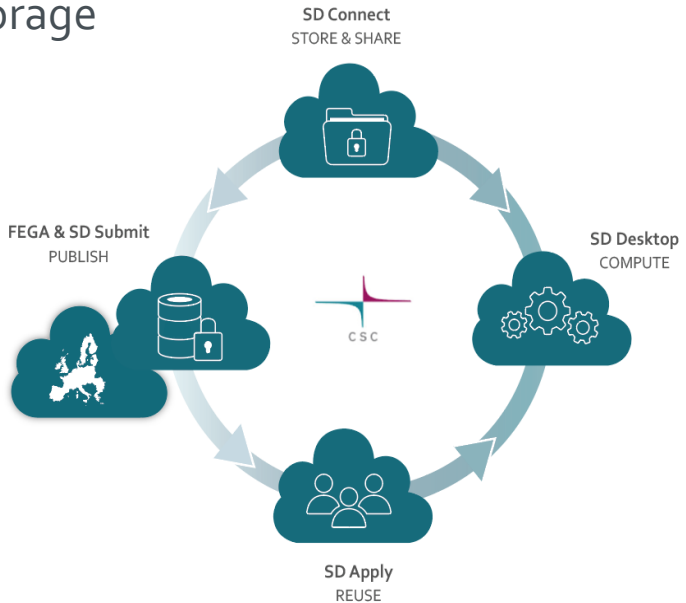
Sensitive Data at CSC

Francesca Morello, Martin Matthiesen, CSC 11.5.2022

Contents

The SD Toolbox

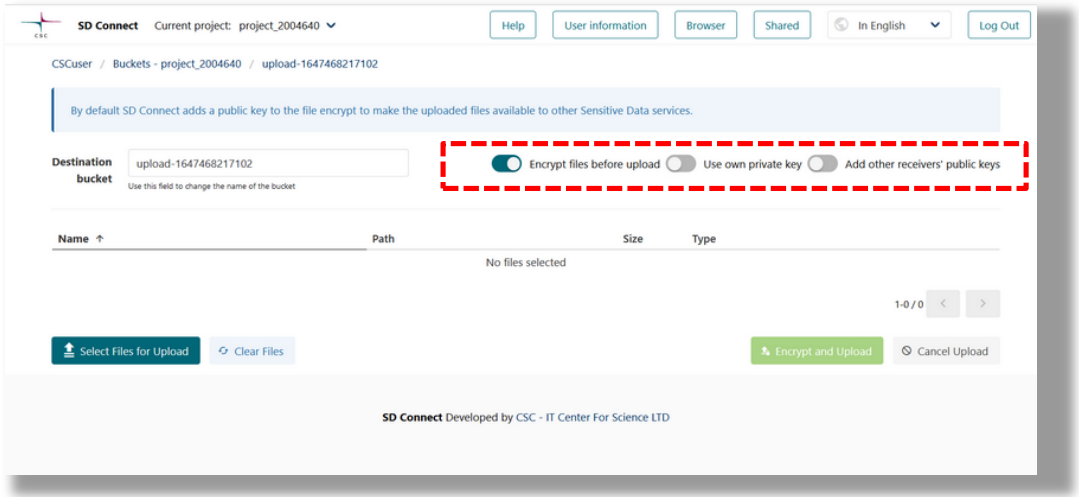
- SD Connect: The encrypted data pipeline and storage
- SD Desktop: A virtual desktop into a closed VM
- SD Submit: The metadata publisher
- SD Apply: links between
  - Data owner
  - Data user
  - Data access committee
- Development funded by Elixir Finland and NeiC
- Operation funded by Finnish Ministry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

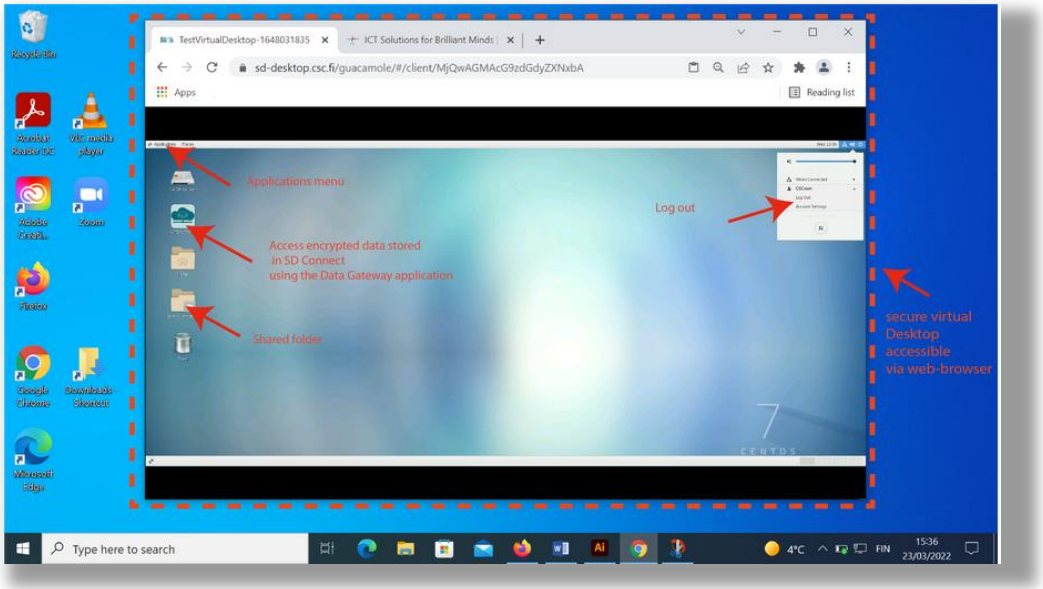


- **Default**: automatic, integrated encryption during upload using the Sensitive Data encryption key
- **Optional:** multiple encryption keys from
  - Researcher
  - Collaborator

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



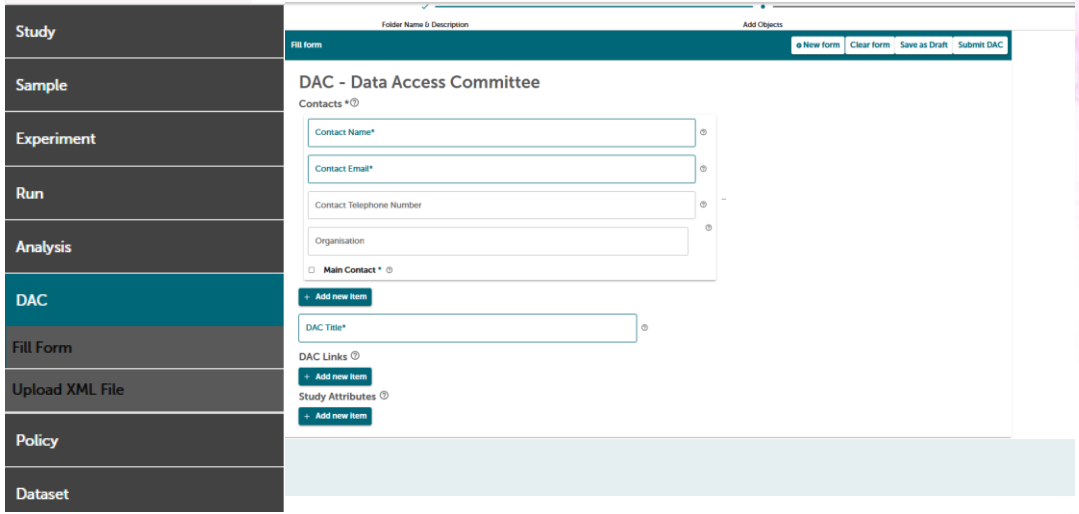
- **Isolated from the internet** (copy/paste disabled)
- Analyse encrypted data stored in SD Connect (via Data Gateway)
- Data export (CSC project manager): results can be **exported** using a specific application (airlock, only CLI)
- Can be **deleted or launched on demand**

SD Submit: publish under controlled acces

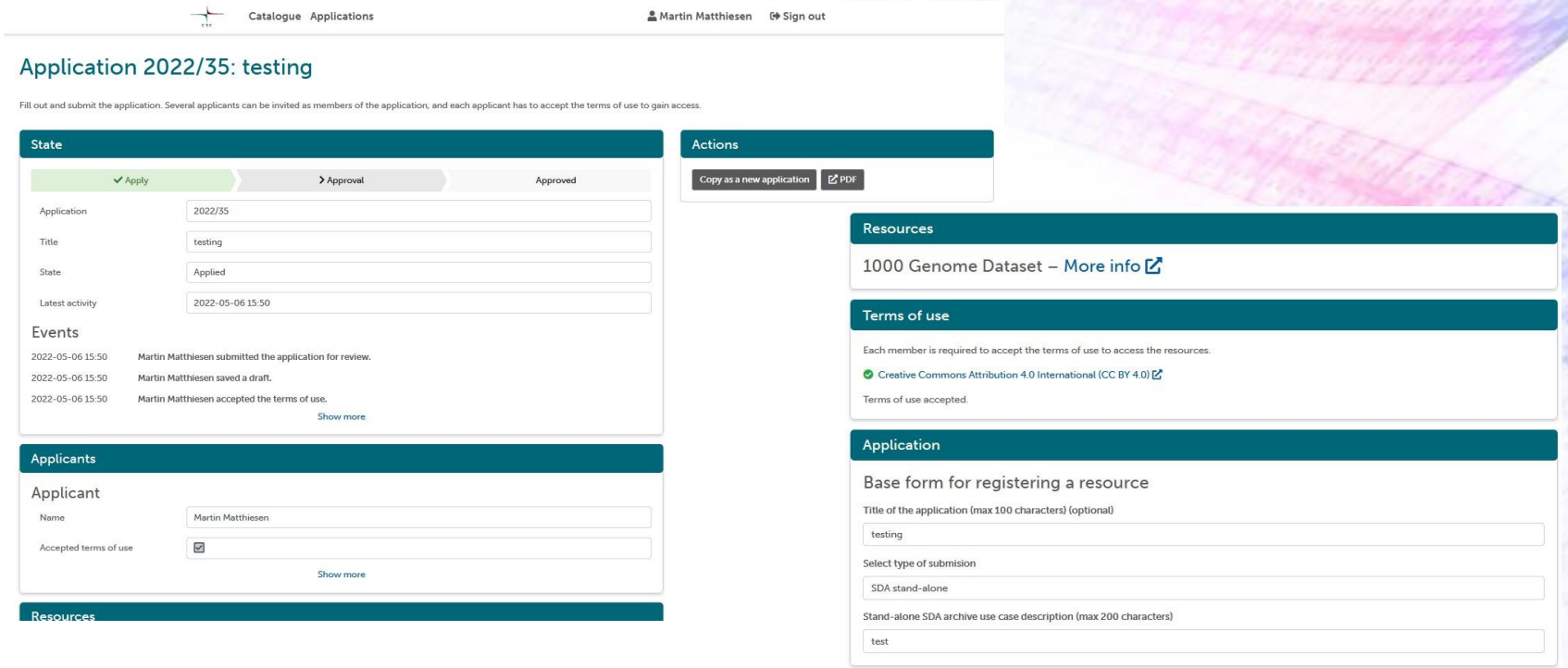
- Dataset is uploaded to CSC programmatically (ssh connection)
  - Automatic datatransfer from SD Connect to SD Submit is WIP
- Dataset is described with appropriate metadata
  - Currently biology is supported best
  - Essential for Discovery & Re-use
- Legal agreements currently not automated
- Each dataset linked to :
  - policies which regulate its re-use, complying with data subject consent
  - Data Access Committee (representing the Data Owner)

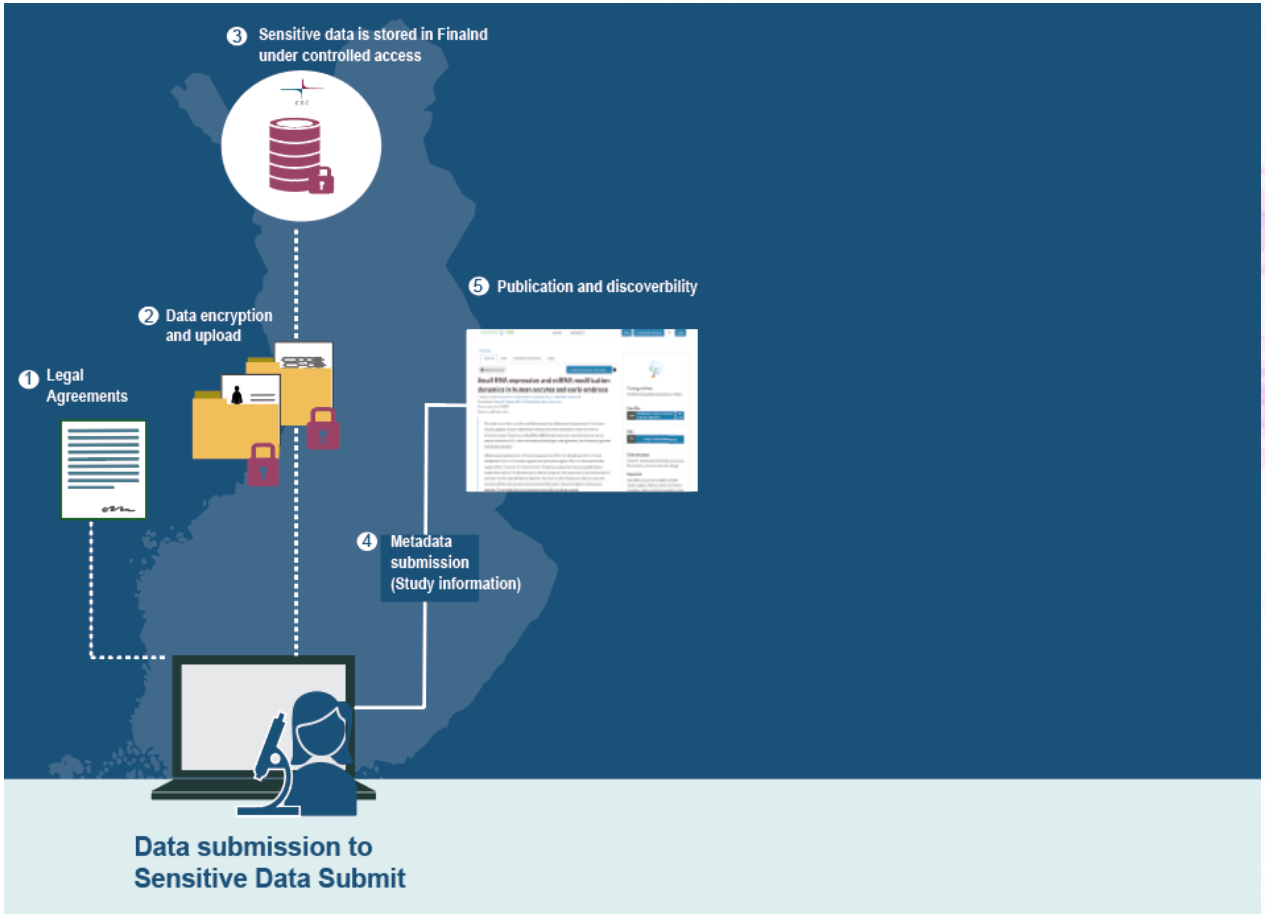
SD Submit: DAC metadata object

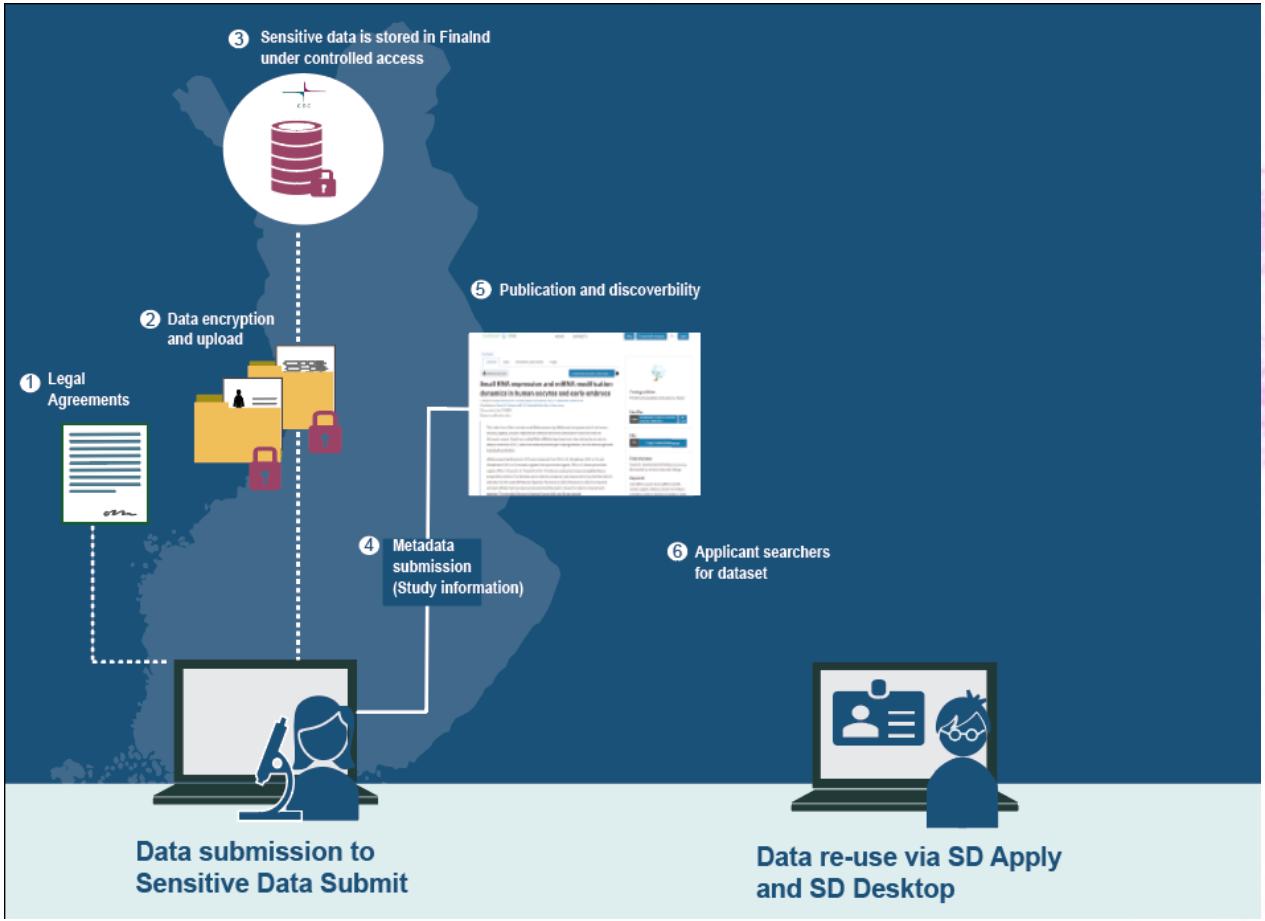
- Link data to a the Data Access Committee that will decide who can access them

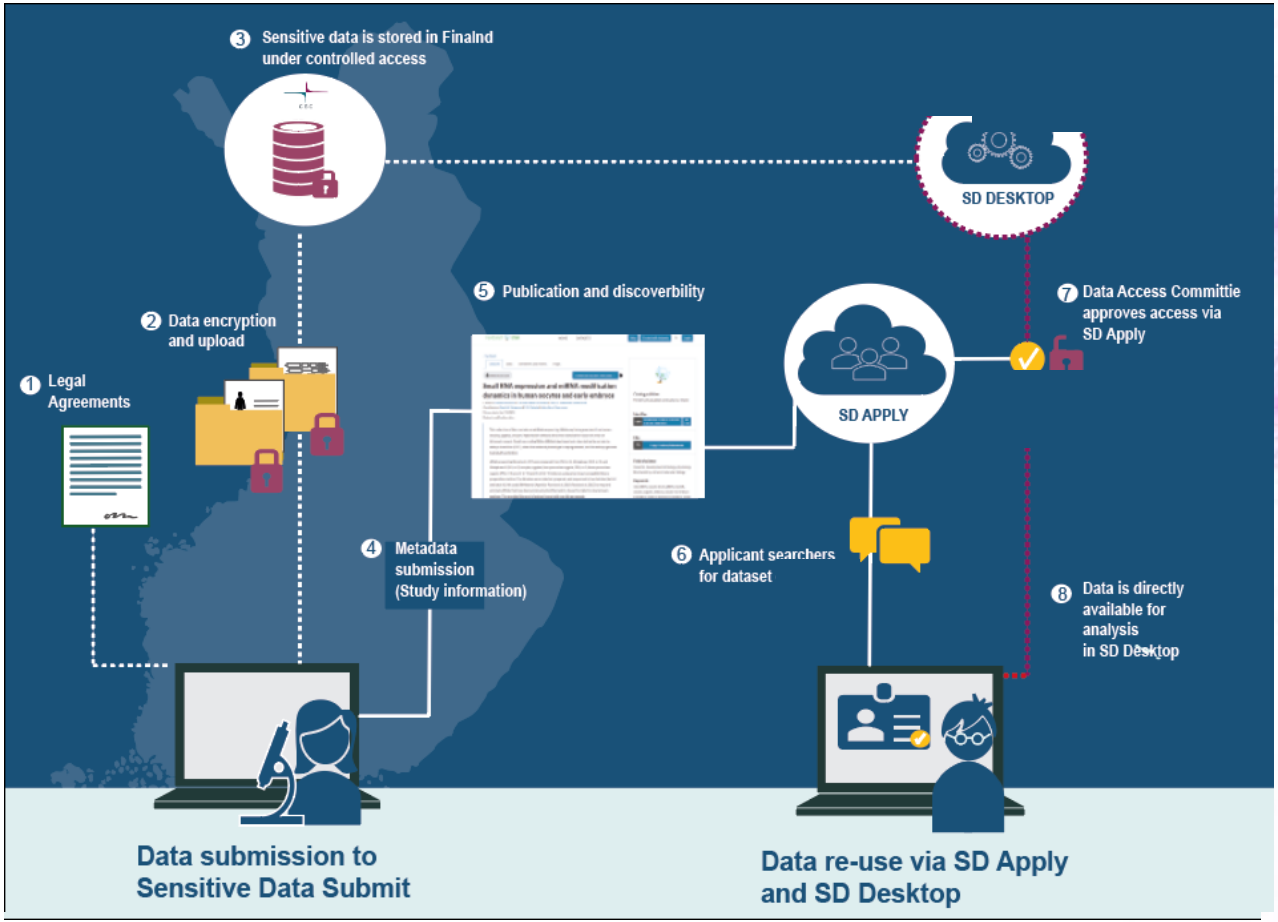


SD Apply: Requesting access




3 Sensitive data is stored in FinaInd under controlled access
5 Publication and discoverbility
2 Data encryption and upload
1 Legal Agreements
4 Metadata submission (Study information)
Data submission to Sensitive Data Submit


3 Sensitive data is stored in Finalnd under controlled access
2 Data encryption and upload
1 Legal Agreements
5 Publication and discoverbility
4 Metadata submission (Study information)
6 Applicant searchers for dataset
Data submission to Sensitive Data Submit
Data re-use via SD Apply and SD Desktop


3 Sensitive data is stored in Finalnd under controlled access
SD DESKTOP
5 Publication and discoverbility
7 Data Access Committie approves access via SD Apply
2 Data encryption and upload
1 Legal Agreements
SD APPLY
4 Metadata submission (Study information)
6 Applicant searchers for dataset
8 Data is directly available for analysis in SD Desktop
Data submission to Sensitive Data Submit
Data re-use via SD Apply and SD Desktop

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