CSC

ICT Solutions for
Brilliant Minds

The trouble with sensitive data

Heikki Lehväslaiho 2024-09-05

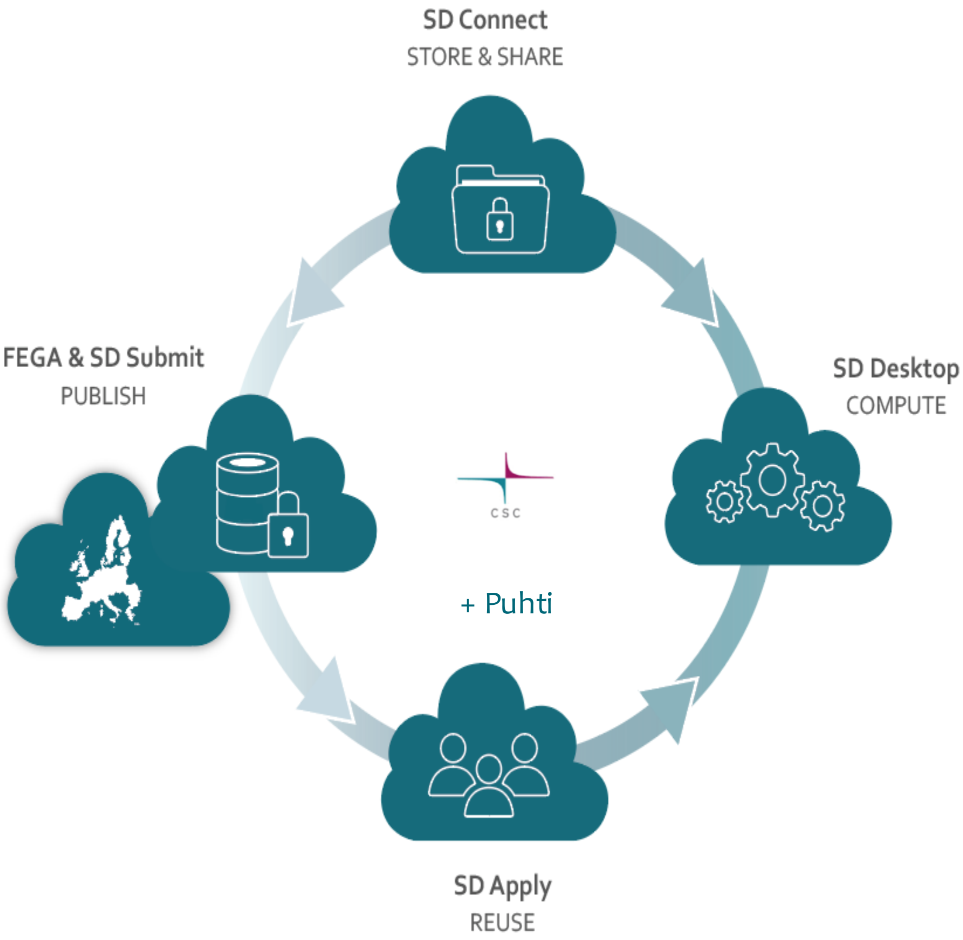
CSC – IT Center for Science

- National scientific service center for scientific computing in Finland
- A non-profit company owned by the Ministy of Education and Culture, and Finnish universities
- Provides:
    - the academic FUNET network, supercomputing and wide variety of generic and specialised computing services
    - the Finnish ELIXIR node for life sciences
    - Sensitive Data Management unit

Approaches to sensitive data processing at CSC

1. Isolated compute environment
    - ePOUTA for academic organisations
        - Infrastructure as a Service (IaaS)
2. Hardened virtual computer environment
    - FIONA for Statistics Finland since 2014
    - KAPSELI for Findata since 2019
        - VIKSU for Findata internal use
3. Sensitive Data life cycle management
    - Sensitive Data (SD) service family
        - SD Connect, SD Desktop, SD Submit, SD Apply

SD Services overview



SD Connect
STORE & SHARE

SD Desktop
COMPUTE

SD Apply
REUSE

FEGA & SD Submit
PUBLISH

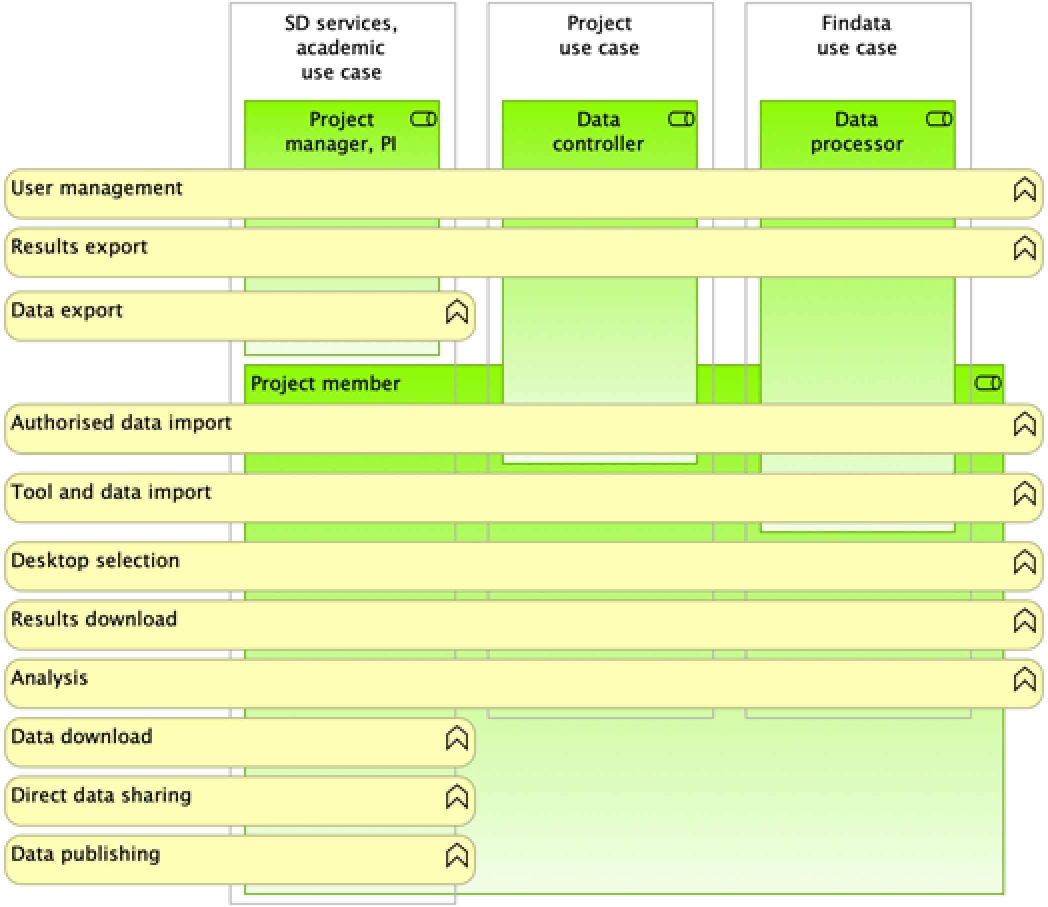
+ Puhti

Different use cases

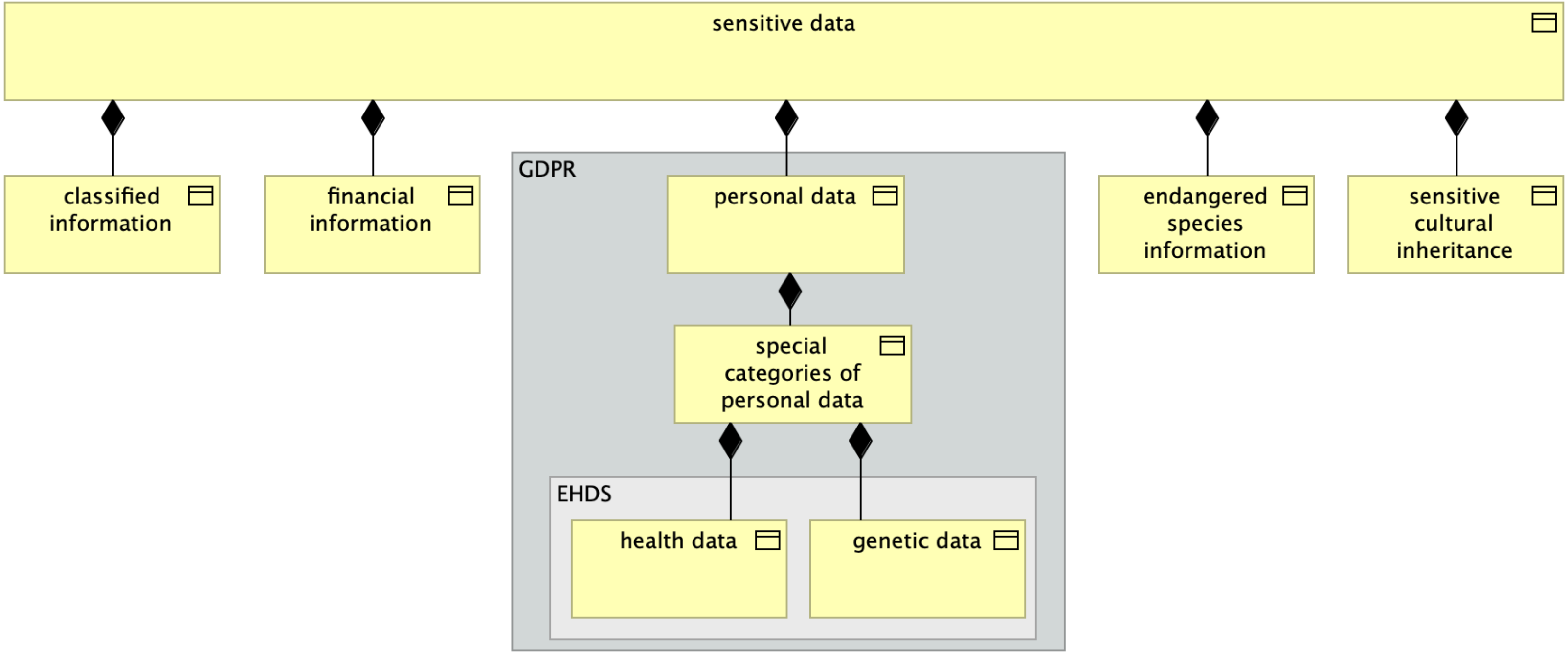
Funded by
the European Union

SD use cases

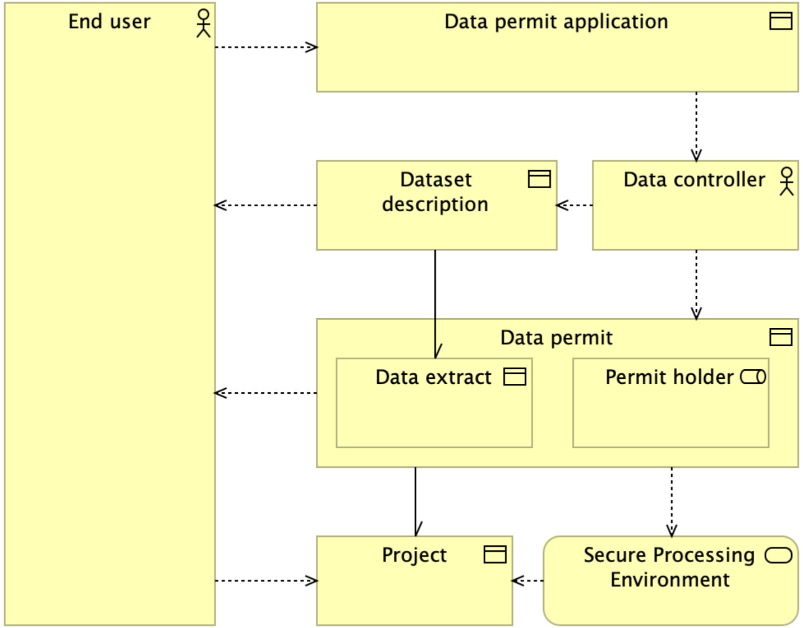
- Functions available to roles depend on the legal basis of the use case
- The project use case is single registry use case under the secondary use law that does not involve Findata



Sensitive data concept model

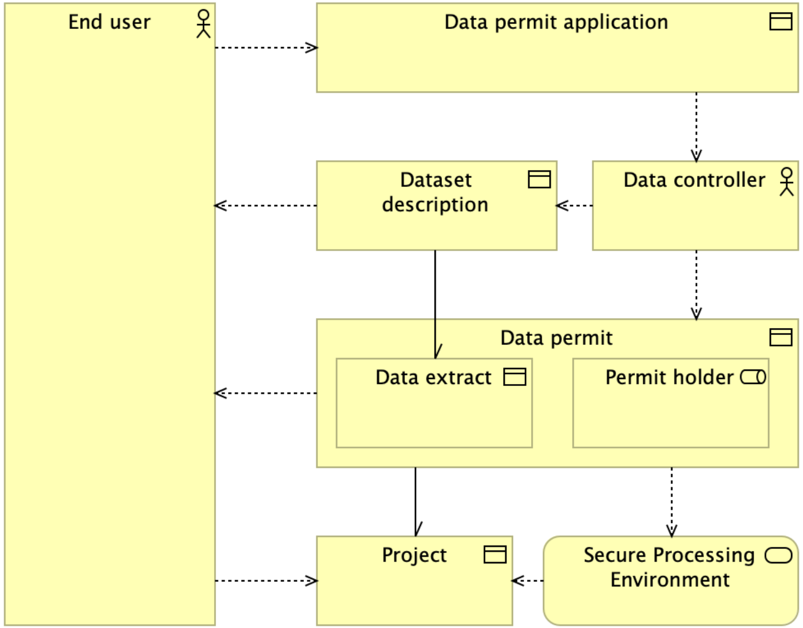


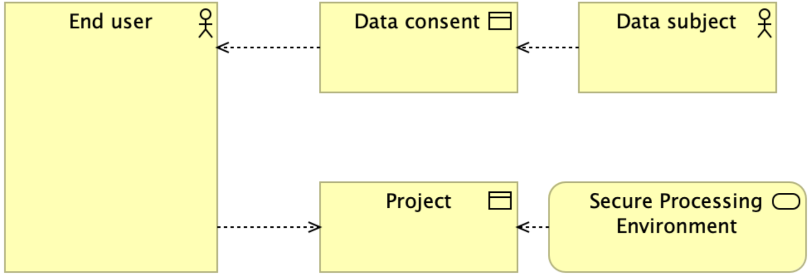
Secondary use of health data



Secondary use of health data

Primary use of research data





SD Services maxims

1. Promote collaboration within the project, prevent information escaping out of it.

2. Trust the users but make sure they are accountable for their actions.

3. Follow the Unix philosophy for minimalist, modular software development.

Principles

Funded by
the European Union

Sensitive data processing principles

1. Enable
   - 1.1. FAIR
     - 1.1.1. Findable
     - 1.1.2. Accessible
     - 1.1.3. Interoperable
     - 1.1.4. Reusable
   - 1.2. Scalable
     - 1.2.1. Flexible

2. Trust
   - ○ 2.1. Security
   - ○ 2.2. Privacy
   - ○ 2.3. Transparency

The priority of problems

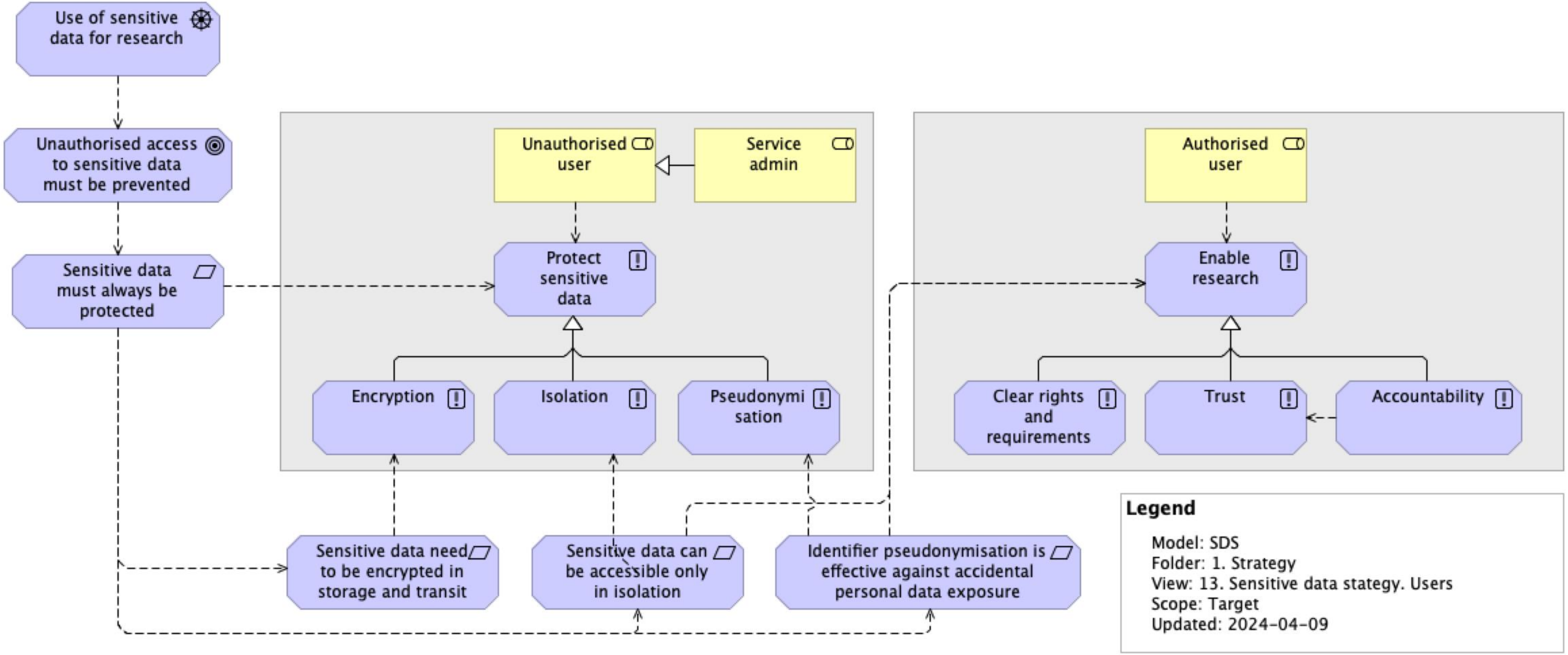
1. Data protection
2. Functionality
3. Security
4. Technicalities

Sensitive data protection requirements

1. Sensitive data MUST be made available in unprotected format only to authorised users
2. Sensitive data MUST be in protected format at rest and in transit
3. Sensitive data protection MUST be done with widely accepted, secure algorithms combined with effective isolation measures

- Protected format ~ currently: encrypted

Example of a simplified logic chain


Use of sensitive data for research
Unauthorised access to sensitive data must be prevented
Sensitive data must always be protected
Unauthorised user
Service admin
Protect sensitive data
Encryption
Isolation
Pseudonymi sation
Authorised user
Enable research
Clear rights and requirements
Trust
Accountability
Sensitive data need to be encrypted in storage and transit
Sensitive data can be accessible only in isolation
Identifier pseudonymisation is effective against accidental personal data exposure

Legend

Model: SDS
Folder: 1. Strategy
View: 13. Sensitive data stategy. Users
Scope: Target
Updated: 2024-04-09

Data Governance Act 2020 Article 2 (14)

‘secure processing environment’ means the physical or virtual environment and organisational means to provide the opportunity to re-use data in a manner that allows for the operator of the secure processing environment to determine and supervise all data processing actions, including to display, storage, download, export of the data and calculation of derivative data through computational algorithms.

EHDS Article 50 (compromise)

1. The health data access bodies shall provide access to electronic health data ***pursuant to a data permit*** only through a secure processing environment, with technical and organisational measures and security and interoperability requirements. In particular, ***the secure processing environment*** shall ***comply with*** the following security measures:

(a) restrict access to the secure processing environment to authorised ***natural*** persons listed in the respective data permit;

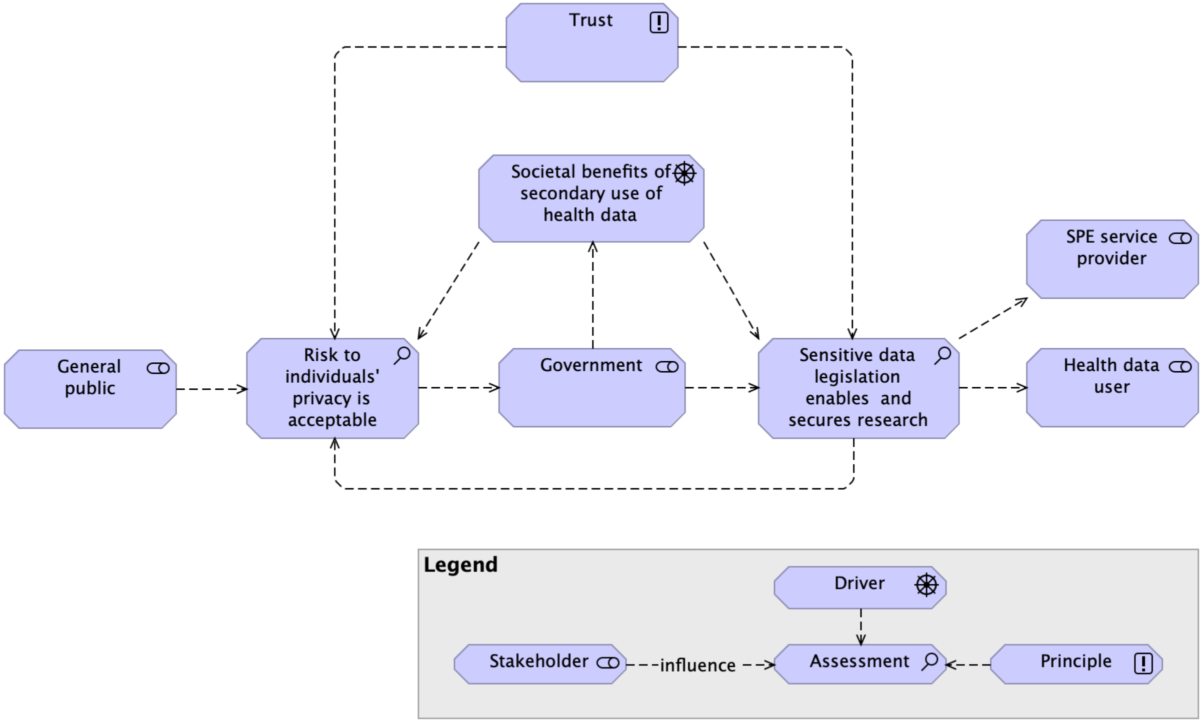
…

(e) keep identifiable logs of access to ***and activities in*** the secure processing environment for the period of time necessary to verify and audit all processing operations in that environment. ***Logs of access should be kept for not shorter than one year;***

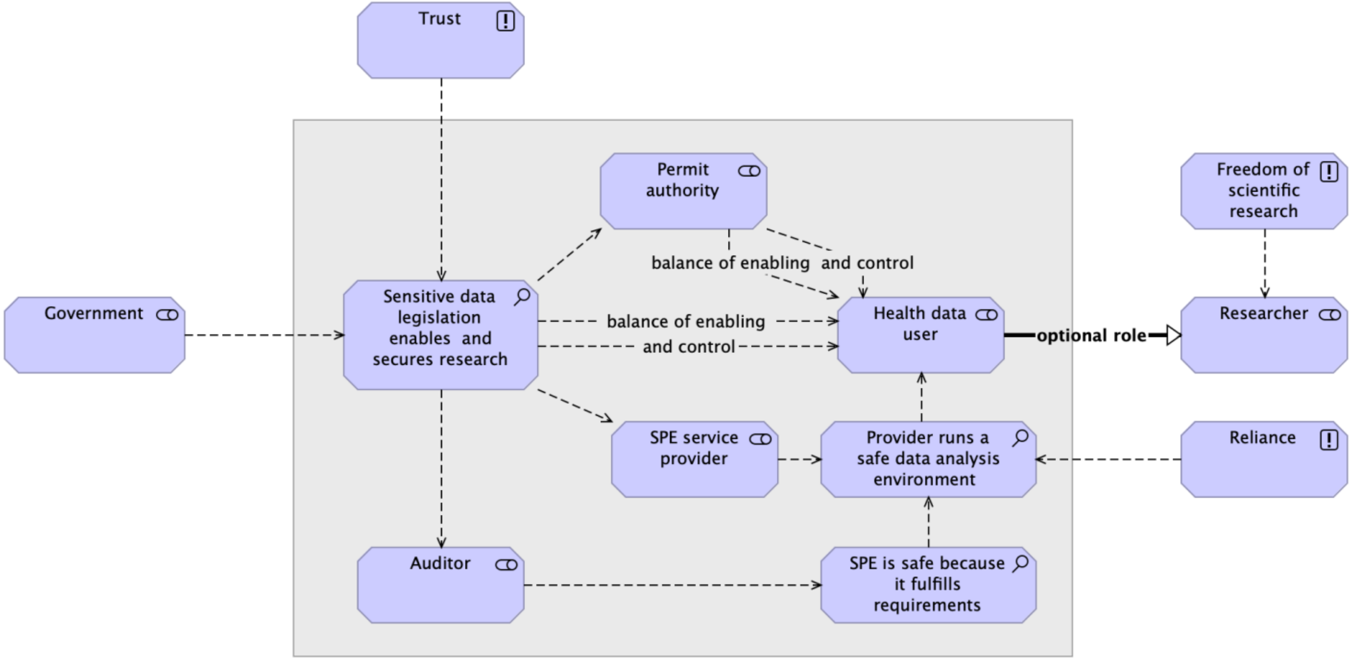
Trust

Funded by the European Union

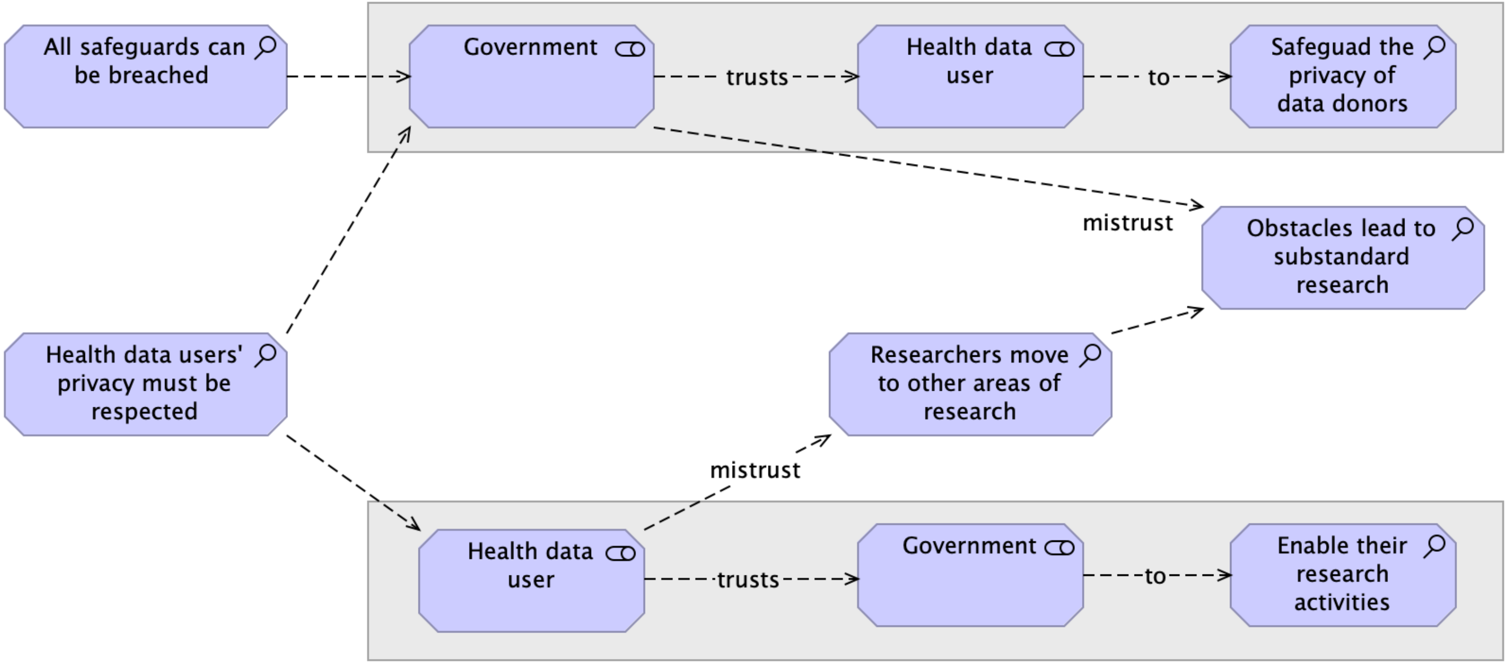
Trust in secondary use of health data 1.1


Trust
Societal benefits of secondary use of health data
SPE service provider
General public
Risk to individuals' privacy is acceptable
Government
Sensitive data legislation enables and secures research
Health data user
Legend
Driver
Stakeholder
influence
Assessment
Principle

Trust in secondary use of health data 1.2


Trust
Government
Sensitive data legislation enables and secures research
Permit authority
balance of enabling and control
Health data user
balance of enabling and control
optional role
Freedom of scientific research
Researcher
SPE service provider
Provider runs a safe data analysis environment
Reliance
Auditor
SPE is safe because it fulfills requirements

Mutual trust in sensitive data research


All safeguards can be breached
Government
trusts
Health data user
to
Safegual the privacy of data donors
mistrust
Obstacles lead to substandard research
Health data users' privacy must be respected
Researchers move to other areas of research
mistrust
Health data user
trusts
Government
to
Enable their research activities

Work in progress …

Funded by
the European Union

Current challenges

- Interoperability
  - Enabling HPC and federated computing
- Limits to users to import and install their own software create problems to scientific research and scalability of service (SaaS vs. PaaS)
- Export of personal information
  - Scientific processes must enrich existing knowledge
- SPE use beyond sensitive data end user processing
  - Transient dataset management
  - Sensitive data infrastructure data processing

Important ongoing projects on sensitive data processing

- EOSC ENTRUST
  - A community-led Europe-wide requirements for interoperability for sensitive data environments
- TEHDAS2
  - Recommendations for EHDS implementation acts

CSC

facebook.com/CSCfi

twitter.com/CSCfi

youtube.com/CSCfi

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

github.com/CSCfi

Kuvat CSC:n arkisto, Adobe Stock ja Thinkstock