Sensitive Data from a Nordic to a EOSC Perspective

(a back and forth journey)

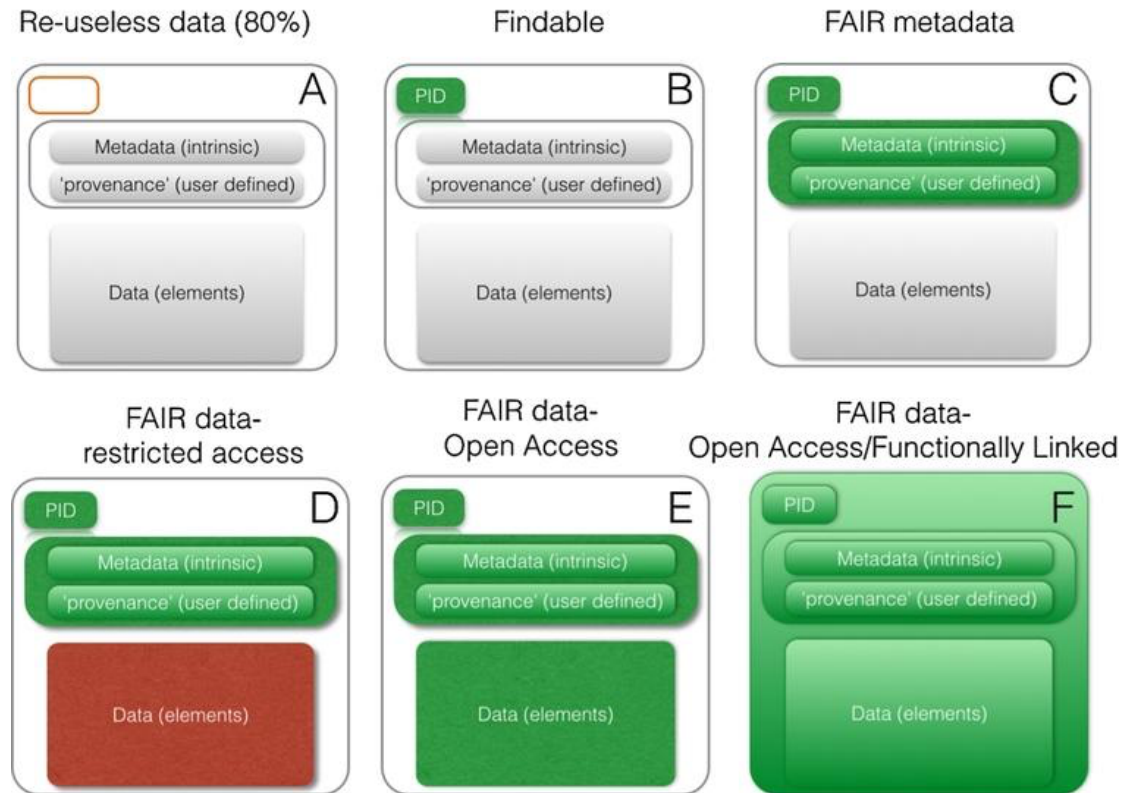**NEIC2019, Copenhavn 14-16 May 2019**

Open Science and Sensitive Data

- The **Open Science Vision**: The possibility to combine dataset from different disciplines/locations and enable compute-intensive analysis (artificial intelligence, machine learning) in secure data driven e-Infrastructure, greatly enhances the scientific research value (es.: personalised medicine, rare-diseases etc…)

- The **Sensitive Data Challenge**: Difficult to move data away from the custodian location (technical issues, legal issues)

FAIR and Sensitive Data

- **Move the tools, not the data**
- Attach rich metadata
  - Anonymised where needed
  - Gathered at central catalogue
- Verifiable research
  - Can we keep track of provenance?
  - advantage of controlled "executables"

Data as increasingly FAIR Digital Objects

Re-useless data (80%) · Findable · FAIR metadata

A · B · C

PID · PID

Metadata (intrinsic)
'provenance' (user defined)

Data (elements)

FAIR data-restricted access · FAIR data-Open Access · FAIR data-Open Access/Functionally Linked

D · E · F

PID · PID · PID

Metadata (intrinsic)
'provenance' (user defined)

Data (elements)
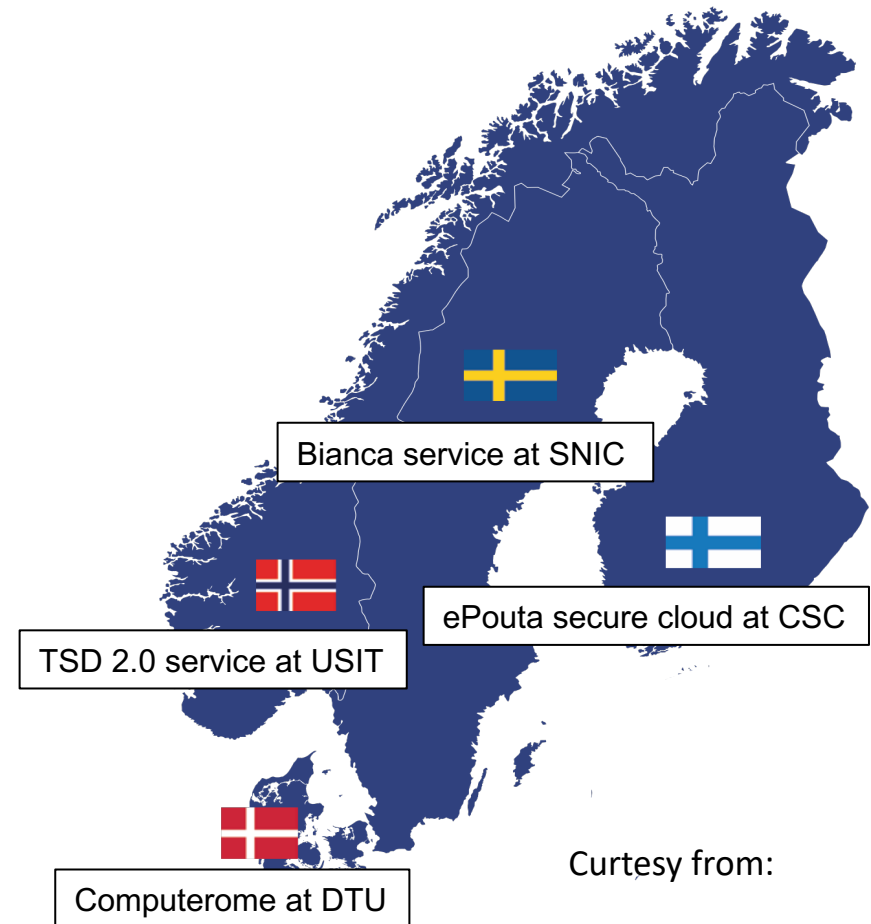
Courtesy of Barend Mons, GoFAIR

Sensitive Data in the Nordic

- Similar legislations with regard to health and personal privacy

- Very close, geographically and economically

- Long history of technical and scientific collaboration

Nordic countries as a sand box to experiment solution for cross-boarders research with sensitive data

Sensitive Data in the Nordic

- 2012-2014 : National secure servers for sensitive data in research were first relised - TSD in Norway, Computerome in Danmark, ePouta in Finland, Mosler in Sweden
  - In 2017 Bianca in Sweden was set in production

- 2014 - : NEIC/Tryggve projects to support development and facilitate access to the sensitive data services in the region for large scale, cross border biomedical research

Bianca service at SNIC

ePouta secure cloud at CSC

TSD 2.0 service at USIT

Computerome at DTU

Curtesy from:

neic

Tryggve

Open Science and Sensitive Data

- 2018 - : EOSC-hub is financed by the EC

- EOSC-hub Mission: EOSC-hub brings together multiple service providers to create the Hub: a single contact point for European researchers and innovators to discover, access, use and reuse a broad spectrum of resources for advanced data-driven research

- Sensitive Data Activities in the EOSC-hub:
  - WP6.6 produce solutions to enhance interoperability of sensitive data services
  - ELIXIR Competence Center
  - WP2 ---> you will learn more in a moment!

EOSC-hub

TSD and ePouta in the Marketplace

EUROPEAN OPEN SCIENCE CLOUD MARKETPLACE

Find service...   All services   🔍   My EOSC Marketplace ▾

⌂ › Services › Storage › TSD

TSD
Services for sensitive data

Provided by:
University of Oslo (UiO)

Research area:
Interdisciplinary

Dedicatedor:
Researchers

☆☆☆☆☆ (0.0 /5)   0 reviews

Order

Want to ask a question about this service?

EUROPEAN OPEN SCIENCE CLOUD MARKETPLACE

Find service...   All services   🔍   🔒 Login   PLACES AND LANGUAGES

Europe

English

⌂ › Services › Compute › CSC ePouta

CSC ePouta
Secure and cost-effective cloud computing for processing sensitive data

Provided by:
CSC

Research area:
Interdisciplinary, Social Sciences, Arts, Philosophy, ethics and religion, Electrical, electronic and information engineering

Dedicatedor:
Researchers, Research organisations, Research group, Providers
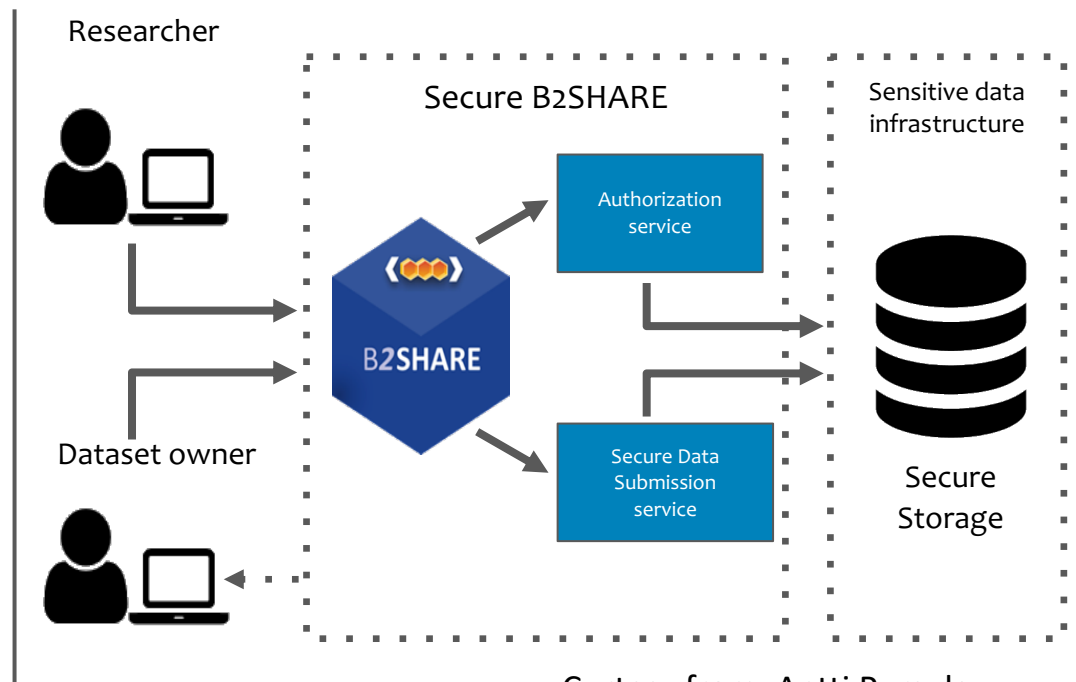
☆☆☆☆☆ (0.0 /5)   0 reviews

Order

If you want to ask a question about this service please login

C S C

WP6.6 - ePouta-B2SHARE integration

- Work in progress to produce **Secure B2SHARE** environment at CSC
- Target for Secure B2SHARE:

  ➢ Store sensitive data & make non-sensitive metadata searchable

  ➢ Controlled access based on permissions decided by the data owner

Curtesy from: Antti Pursula

- Data anonymization tool, that allows to remove identifying information from data.

ID	Age	Zipcode	Diagnosis
1	28	13053	Heart Disease
2	29	13068	Heart Disease
3	21	13068	Viral Infection
4	23	13053	Viral Infection
5	50	14853	Cancer
6	55	14853	Heart Disease
7	47	14850	Viral Infection
8	49	14850	Viral Infection
9	31	13053	Cancer
10	37	13053	Cancer
11	36	13222	Cancer
12	35	13068	Cancer

k-anonymization

ID	Age	Zipcode	Diagnosis
1	[20-30]	130**	Heart Disease
2	[20-30]	130**	Heart Disease
3	[20-30]	130**	Viral Infection
4	[20-30]	130**	Viral Infection
5	[40-60]	148**	Cancer
6	[40-60]	148**	Heart Disease
7	[40-60]	148**	Viral Infection
8	[40-60]	148**	Viral Infection
9	[30-40]	13***	Cancer
10	[30-40]	13***	Cancer
11	[30-40]	13***	Cancer
12	[30-40]	13***	Cancer

OpenAIRE

amnesia

9

Curtesy from: Azab Abdulrahman

Sensitive Data in the EOSC: are we ready?

- Very little money in the EOSC-hub ☹

- GDPR is in place but…. Still national statutary acts quite different, data cannot move, little trust!

- Lack of trustworthy mechanism to do authorization and vetting

We go back to the Nordics!

National policy programs for Integrated Health Data

In the Nordic countries, governments has been developed or are currently developing strategies to produce integrated national solutions for storing, discover, re-use, aggregate health data from different registries, health care institutions and data producers.

- In many of the Nordic countries the existence of a unique personal identifier allows cross-linking of personal data though different sources. This puts the Nordic countries ahead in building integrated health data systems as compared to other regions.

From Norsk Helse Analyse Platformen
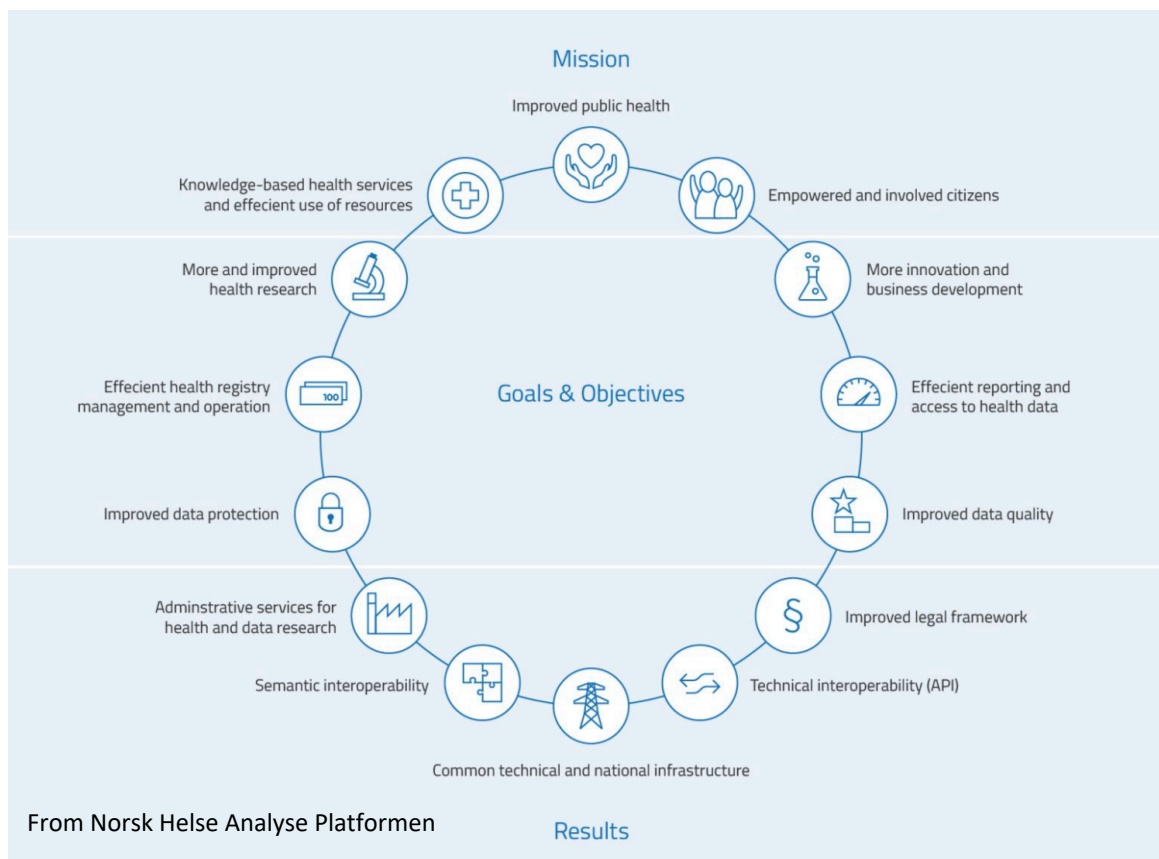
Nordic Program of Health and Welfare

In 2018 a Nordic Secure Cloud Working Group was established in the contest of the Nordic Program of Health and Welfare to work at a joint solution proposal for a Nordic Secure Cloud Solution.

→ **more about vision and objectives in the next talk: Dr Juni Palmgren**

NordForsk

| Programmes and projects | Apply for funding | Research policy and |

Nordic biobanks and registers - A basis for innovative research on health and welfare

This report emphasises registers and biobanks as research infrastructures for innovative research on health and welfare. It gives an overview of current knowledge on cross-border development of research that uses Nordic registers, biobanks and clinical studies, and includes suggestions for policy implications.

🛒 ORDER THE PUBLICATION   ⬇ DOWNLOAD

Programme: Nordic Programme on Health and Welfare

Open Science and the Nordic Added Value (Nordisk Nytte)

✓ **Open Science Programmes** are strong in the Nordics and many actors are **already involved in EOSC related initiatives**

✓ There is a good **history of Research & Policy collaboration** across the Nordics upon which we can build

✓ As individual countries, the Nordic countries are probably too small to have a significant influence on the EOSC. However, put together, they reach a **critical mass** which can strengthen their position

✓ By coordinating their action, the Nordic countries can **move faster on certain strategic areas** and **influence** the EOSC developments & agenda, as well as their collaborative culture & values

Dr. Damien Lecarpentier

Call: INFRAEOSC-05-2018-2019
Partners: 24
Budget: 5.9M€
Coordinator: Gudmund Høst, NeIC
Director
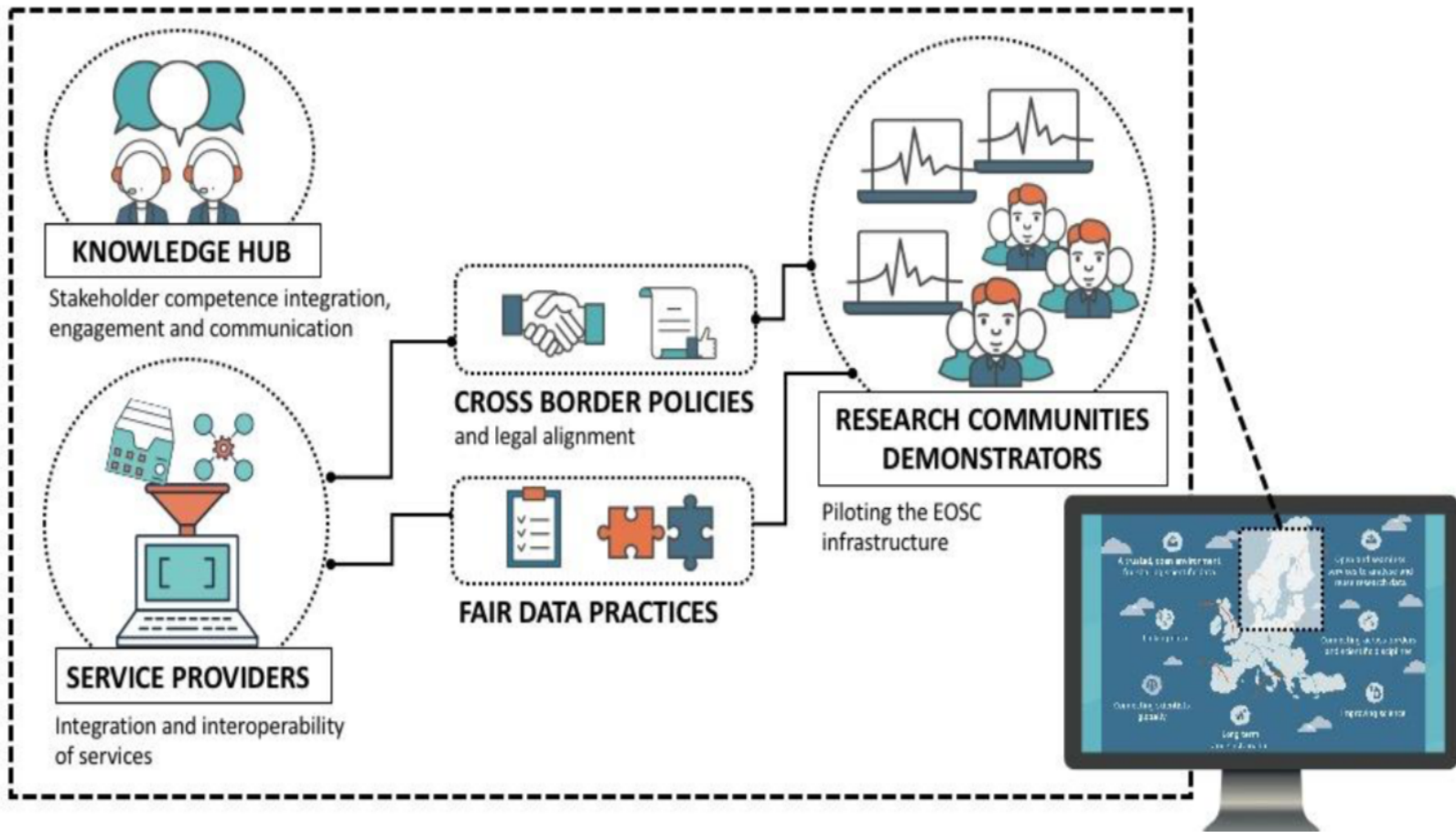
Project period: Sep, 2019 – Aug, 2022
Kick off mtg: Sep 2-3, 2019 (CSC, FI)

EOSC Nordic

DEMONSTRATING THE POTENTIAL OF EOSC

USING THE NORDIC AND BALTIC COUNTRIES AS A TESTBED ENVIRONMENT.

DISCOVER AND RE-USE RESEARCH DATA
Foster the discoverability and re-use of research data. Harvesting metadata in the future EOSC **metadata catalogue**.

ANALYSIS AND POST-PROCESSING
Integration of community specific portals with large scale computing facilities and pilot usage of computing resources across borders

SENSITIVE DATA AND ORCHESTRATION
Piloting an orchestration mechanism that brings national secure servers together, allowing the analysis of sensitive data without moving data away from the custodian.

DATA MANAGEMENT SHARING AND ARCHIVING
Facilitating cross-borders data sharing by enabling data management on a distributed environment.

EOSCpilot
The European Open Science Cloud for Research Pilot Project

EOSC Nordic: Sensitive data and orchestration

- **Goal**: facilitating cross border research involving sensitive data by showcasing how a secure and compliant cloud infrastructure for sensitive data could be provided as part of EOSC

- **High level idea**: to ensure that researchers have an easy and secure platform to work with large amounts of sensitive data under uniform conditions across the countries without moving the data from their original location

EOSC Nordic: Sensitive data and orchestration

- **What type of data?** The use case stems from communities working on personalized heart medicine but the solution will be applicable by design to any research community in the field of medical science

- **Use Case**: a platform for personalized hearth medicine allowing analysis combining the local Danisk repository data with data from remote repositories located in other countries. Use case adopted also to support the vision developed in the technical cloud working group in the framework of the Nordic Commons under the Nordic Program on Health and Welfare.

EOSC Nordic: Sensitive data and orchestration

- Link to **NEIC/Tryggve** project – use cases, collaborations

- Link to the **Nordic Common** – vision, use case

"Nordic Trust Region"?

- **What if the data need to be moved?**

- **What would technically facilitate the establishment of a Nordic Trust Region**?

  - Cross borders mechanism to do strong identification of the respondents
  - Federated framework
  - Homogenous conditions for collecting consent and supporting opt-out decisions
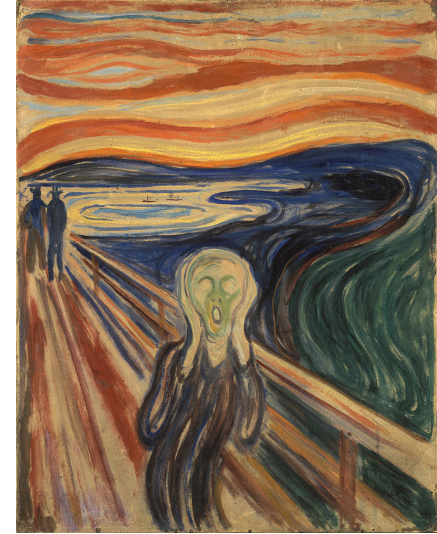
EOSC and sensitive data: a back and forth journey to the mars or to the moon?