



Infrastructure for research with sensitive data

RDA Finland Workshop 25.4.2019

Antti Pursula, CSC



Contents

- Challenge for infrastructure for research with sensitive data
- CSC's national solutions for managing sensitive data
- European and Nordic collaboration to support cross-border research utilizing sensitive data

Research with personal and/or sensitive data

Many research topics (such as health and genetics, and social sciences) involve the use of personal data in order to reach scientific goals.

Society benefits when such data collections are made available for research

Privacy of the individuals need to be protected, on the other hand, so that personal data is used only for approved purpose and under high technological safeguards.

Privacy and integrity of individuals need to be protected, thus limiting the use of sensitive data

Infrastructure for research with sensitive data

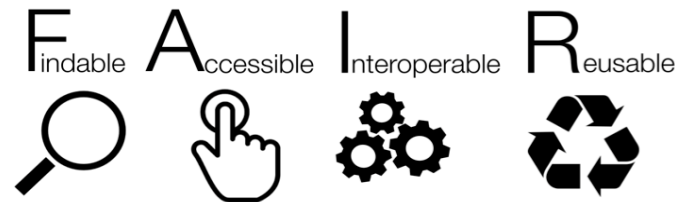
Challenge is to utilize personal data in research for the benefit of society and individuals, while preventing non-authorized access to it.

Infrastructure for research with sensitive data

Challenge is to utilize personal data in research for the benefit of society and individuals, while preventing non-authorized access to it.

- Secure IT is needed to solve this challenge
- Requires compliance with legislation and regulations
- FAIR principles in sensitive data as well

Open science with sensitive data



- Trends on FAIR data and open science
- Personal data is **not** open data, due to ethical considerations and legal regulations.
- But personal data used in research can and should be FAIR
 - Requires emphasis on research data management
 - Strict accordance to consents, ELSI conditions, access policies etc.
 - E.g. **Findable** metadata, authorized **Access** in secure environment, **Interoperable** data and metadata formats, **Reusable** under ELSI conditions within secure systems

Sensitive data solutions at CSC



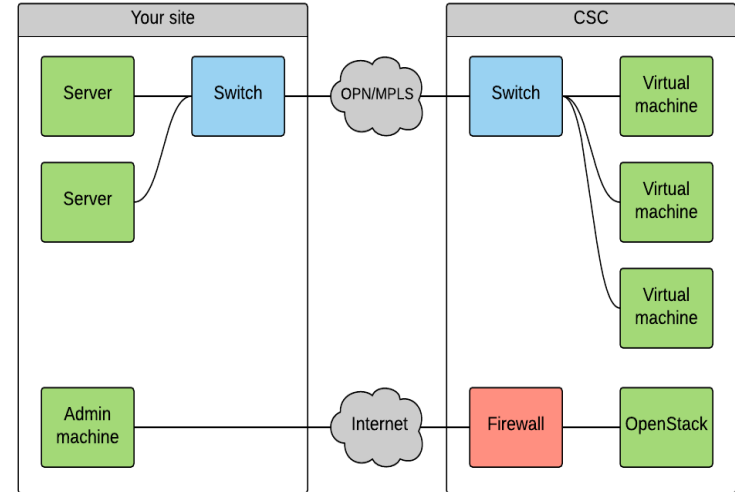
Operational services for sensitive data – CSC ePouta secure cloud



<https://research.csc.fi/cloud-computing>

CSC ePouta – sensitive data processing service

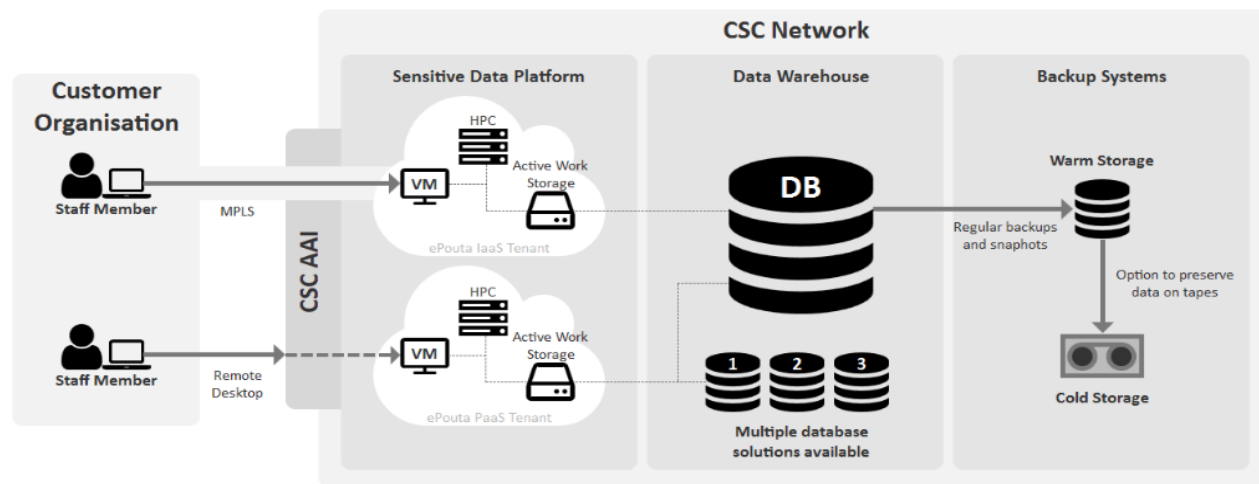
- CSC ePouta secure cloud is a closed environment designed for processing **sensitive data**
- Complete isolation of VM instance, the rest of the world, and other ePouta customers
- Virtual machines are in the customer organization's network through secure and private connection
- The design has been demonstrated to work even across countries (Sweden / Finland)



<https://research.csc.fi/cloud-computing>

CSC Sensitive data management platform

- Work in progress to extend sensitive data service offering at CSC to a complete platform with data archiving, access management and processing within the same environment

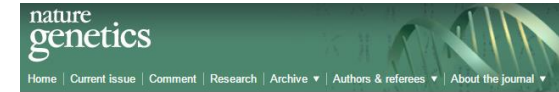


Use cases

- Use cases mainly from biomedical and health area (e.g. ELIXIR, CORBEL, EOSC-Life, Tryggve)
- Other areas include social sciences and humanities, for instance videos from language research

Migraine genome study data of Finns

- Data used by Aarno Palotie and Samuli Ripatti at the Finnish Institute for Molecular Medicine originates from digitalisation of samples of Finns stored in biobanks
- Technical requirements included to transfer large amounts of data for archival in Finland with Nordunet and Funet coming from sequencing centers and databases in the United States
- CSC, Nordunet, FIMM, NIH, UWash, MIT
- <https://doi.org/10.1038/ng.3598>



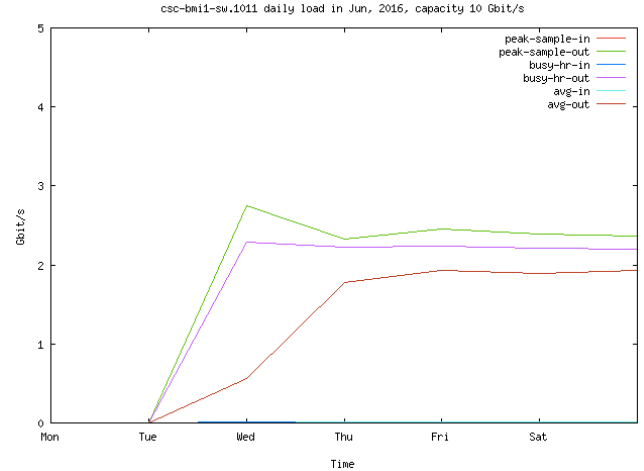
home > archive > issue > article > abstract

ARTICLE PREVIEW
[view full access options >](#)

NATURE GENETICS | ARTICLE

日本語要約

Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine



Statistics Finland remote access system

- Statistics Finland and CSC collaborate in providing secure access to statistics data
- CSC provides secure platform for the FIONA solution
- Data provided to secure platform by Stats Finland after reviewing the requests
- Controlled download of result data

Web site at Statistics Finland: https://www.stat.fi/tup/mikroaineistot/etakaytto_en.html



International collaboration for supporting cross-border sensitive data projects



NeIC Tryggve - Infrastructure for research with sensitive data

Project objective: Tryggve develops and facilitates access to secure e-infrastructure for sensitive data, suitable for hosting **large-scale cross-border biomedical research** studies

- Tryggve develops infrastructure in which secure storage and computing together form a **secure distributed platform** for sensitive data

<https://neic.no/tryggve>

Secure computing and data environments

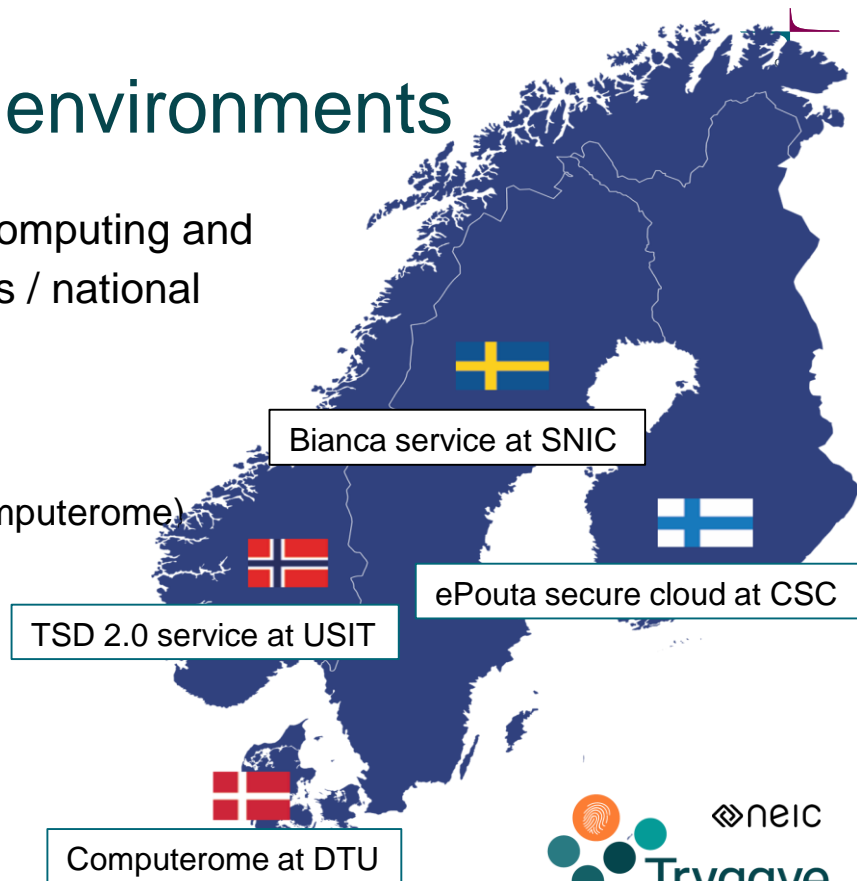
Tryggve service development relies on secure computing and data environments at participating ELIXIR Nodes / national e-infrastructure providers

Secure cloud (IaaS)

- Infrastructure for data and computing (ePouta, Computerome)

Secure remote desktop (PaaS)

- Backed up with computing resources (TSD, Computerome, Bianca, CSC ePouta (to be released))



NeIC Conference

NeIC Conference 14-16 May, 2019, Copenhagen, includes workshop on **”Open Science with Sensitive Data”**.

Registration DL 1st of May.

<https://indico.neic.no/event/18/>



14-16 May 2019
Tivoli Hotel & Congress Center
Europe/Copenhagen timezone

Overview

Timetable

Programme details

Call for Abstracts

Book of Abstracts

Social Activities

Practical Information

Registration

Participation list

Program Committee Chair

✉ lene.krol.andersen@dei...

Open Science with Sensitive Data

📅 15 May 2019, 13:30

🕒 1h 30m

📍 Tivoli Hotel & Congress Center

Open Science

Open Science

Speakers

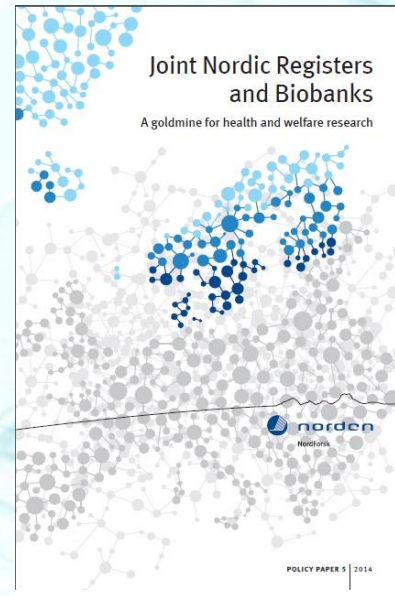
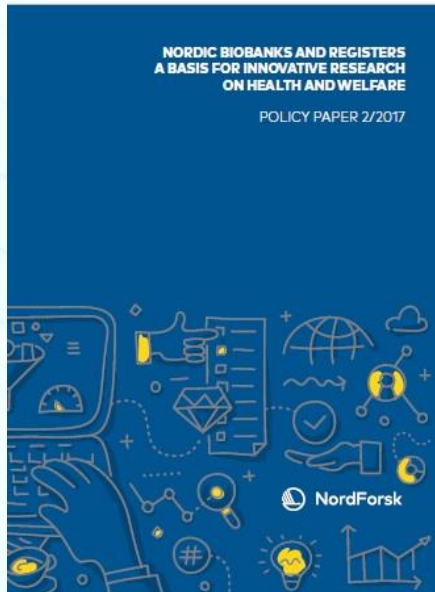
- 👤 Antti Pursula (CSC)
- 👤 Gard Thomassen (USIT, University of O...)
- 👤 Niclas Jareborg (NBIS)
- 👤 Rob Baxter (EPCC)
- 👤 Maria Francesca Iozzi (Sigma2)
- 👤 Juni Palmgren (NordForsk)
- 👤 Jacob Hjelmberg (SDU)
- 👤 Heidi Laine (CSC)

Description

Many scientific fields are using, or would like to use, personal or sensitive data in the research. Such fields include for example genomics, health, social sciences and language research. The sensitive data that has been cleared for secondary use, should be properly managed and made findable under the same principles than non-sensitive research data. This naturally needs to be done under strict ethical and legal compliance and via secure IT services. However, providing secure e-infrastructure for large cross-border research projects dealing with sensitive data is still in great demand and remains in some extents an unsolved challenge. Moreover, the emphasis on open science and FAIR data by the science communities and policy-makers increase the demand for professional research data management in connection to sensitive data.

Further work on Nordic level

- Nordforsk Health and welfare program coordinates an initiative to create Nordic Commons for health data



EUROPEAN OPEN SCIENCE CLOUD

BRINGING TOGETHER CURRENT AND FUTURE DATA INFRASTRUCTURES

A trusted, open environment
for sharing scientific data

Open and seamless
services to analyse and
reuse research data

Linking data

Connecting across borders
and scientific disciplines

Connecting scientists
globally

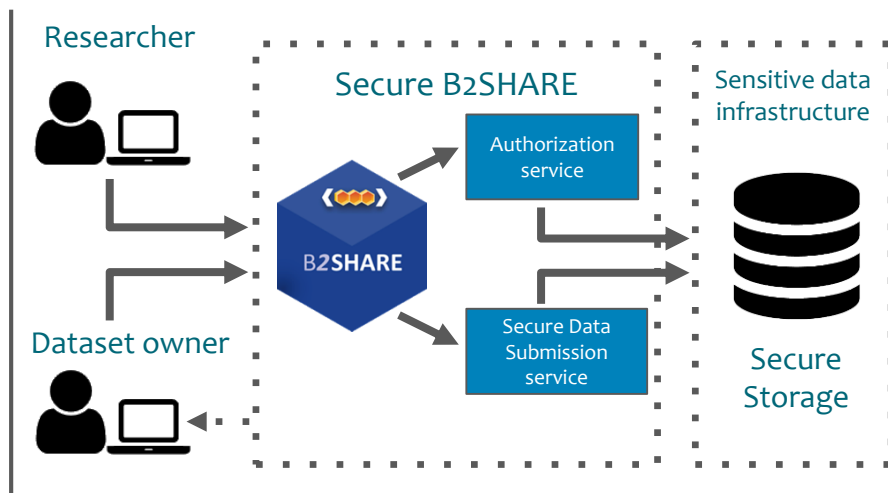
Improving science

Long term
and sustainable



CSC ePouta in EOSC-hub

- CSC is one of the providers of sensitive data services in EOSC-hub
- 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
- [See our poster #18: Secure B2SHARE: Metadata publication and secure sharing of sensitive datasets](#)

Secure services for sensitive data in EOSC-hub



SERVICES ▾

COMMUNITIES ▾

RESOURCES ▾

NEWS

Home » News » Secure services for sensitive data in research

Secure services for sensitive data in research



An overview of secure data services offered through EOSC-hub, by Abdulrahman Azab, Francesca Iozzi and Antti Pursula

Research often involves the use of personal data as a basis for the scientific analysis. However, a particular challenge in this area is to use these data resources without violating privacy. And for that we need secure digital infrastructures, compliant with both national and European regulations.

Article on EOSC-hub Magazine describing sensitive data services in the project

- Online version at

<https://www.eosc-hub.eu/news/secure-services-sensitive-data-research>


CSC ePouta in EOSC Marketplace



**EUROPEAN OPEN
SCIENCE CLOUD
MARKETPLACE**

All services ⇅



 Login

[Home](#) > [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

Dedicated to:

Researchers, Research organisations, Research group, Providers

☆☆☆☆☆ (0.0 / 5) 0 reviews

Order

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

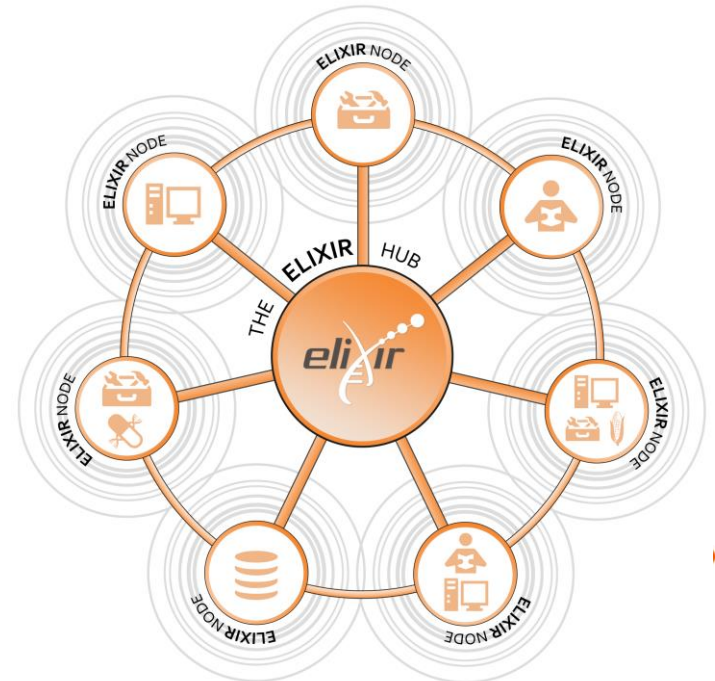
ELIXIR – European research infrastructure for life science data

ELIXIR data infrastructure for Europe's life science research sector

ELIXIR Nodes build local bioinformatics capacity throughout Europe

ELIXIR Nodes build on national strengths and priorities

ELIXIR Finland node is located at CSC, focusing on compute services



ELIXIR Nodes



- ELIXIR Nodes build local bioinformatics capacity throughout Europe
- 23 members, over 200 institutes involved in ELIXIR Nodes
- Sustained international legal framework
- Nodes and national networks have developed own agreements
- CSC operates the ELIXIR Finland node, in collaboration with national biomedical centres



In Finland, Users (e.g. at FIMM and THL) see CSC Compute and Data as a single unified resource



Summary

- Using personal data for research purposes poses challenges for IT infrastructure – but these can be solved!
- CSC ePouta is an operational secure cloud for sensitive research data – ongoing development for comprehensive secure data platform (archive, discover, share, process, reuse)
- Ongoing international work with leading European infrastructures to facilitate sensitive data usage in cross-border projects

Summary

- Using personal data for research purposes poses challenges for IT infrastructure – but these can be solved!
- CSC ePouta is an operational secure cloud for sensitive research data – ongoing development for comprehensive secure data platform (archive, discover, share, process, reuse)
- Ongoing international work with leading European infrastructures to facilitate sensitive data usage in cross-border projects



<https://www.csc.fi/web/blog>

Webinar: What are the policies and possibilities for managing your sensitive data?

<https://www.csc.fi/web/training/-/sensitive-data-webinar-2018>