Nordic platform for sensitive data

Antti Pursula, NeIC
NeIC Tryggve project manager
NeIC 2019 Conference, Copenhagen

The Nordics have unique digital health registers, biobanks, genome and other data collections

Tryggve

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 valuable data collections are made available for research

Privacy of individuals need to be protected, 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

Tryggve - Infrastructure for research with sensitive data

Tryggve is collaboration of NeIC and Nordic ELIXIR Nodes (DK, FI, NO, SE) to develop and provide data and compute services for human data across borders

- NeIC = Nordic e-Infrastructure Collaboration
- ELIXIR = European research infrastructure for life science information

Challenge is to balance benefit for the society and privacy of individuals

- Secure IT is needed in solving this challenge

neic NORDIC E-INFRASTRUCTURE COLLABORATION NordForsk elixir DENMARK elixir FINLAND elixir NORWAY elixir SWEDEN

NeIC Tryggve projects

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

- First project 2014-2017: groundwork, building blocks, piloting
- Current project 2017-2020: larger funding ~6M€, going towards service delivery, increased focus on use cases

https://neic.no/tryggve

neic NORDIC E-INFRASTRUCTURE COLLABORATION NordForsk elixir DENMARK elixir FINLAND elixir NORWAY elixir SWEDEN

Secure computing and data environments

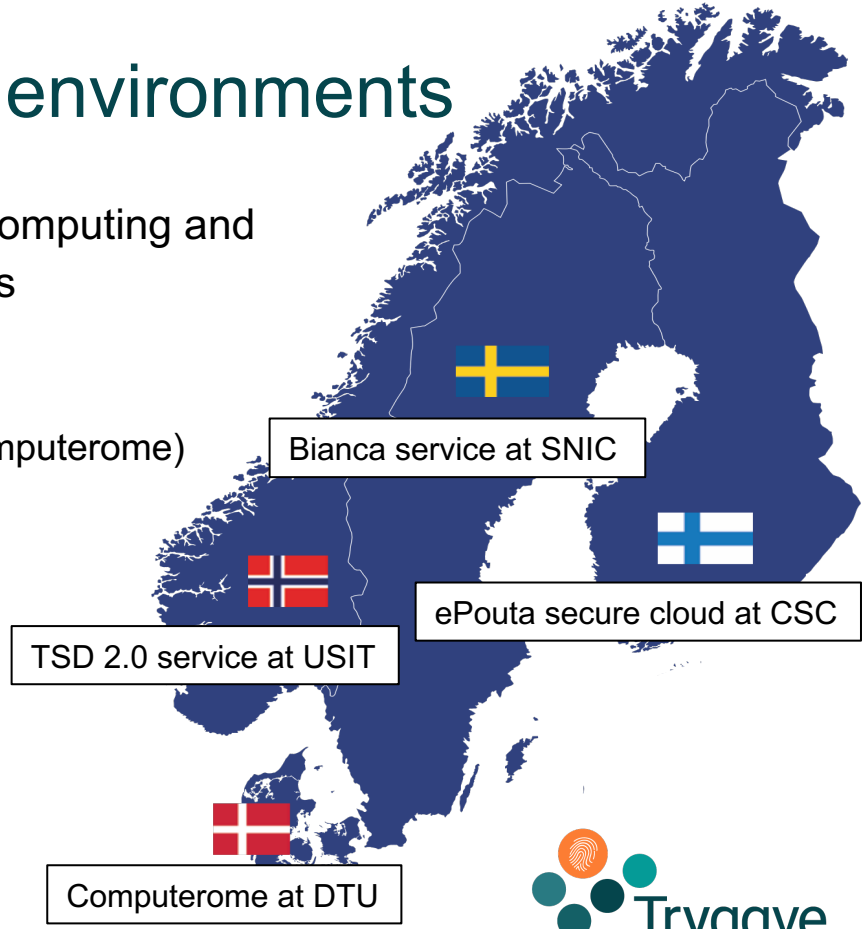
Tryggve service development relies on secure computing and data environments at participating ELIXIR Nodes

Secure cloud (Iaas)

- Infrastructure for data and computing (ePouta, Computerome)

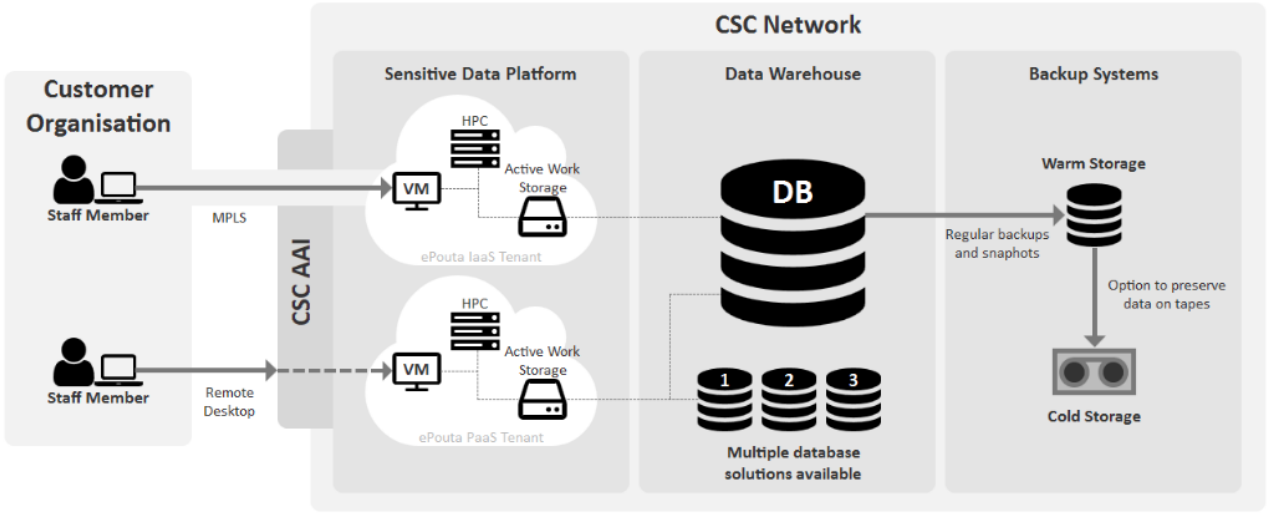
Secure remote desktop (PaaS)

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



Tryggve environment in Finland: CSC ePouta Sensitive data 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



Tryggve activities

- Development of **sensitive data archiving technology**
- Development of **secure tools for analyzing sensitive data across borders**
- Operating a **use case program**
- Targeted development of the secure Tryggve platforms
- Implementing ELIXIR AAI based authentication and authorization solutions
- Providing assistance with GDPR related issues

Tryggve

neic

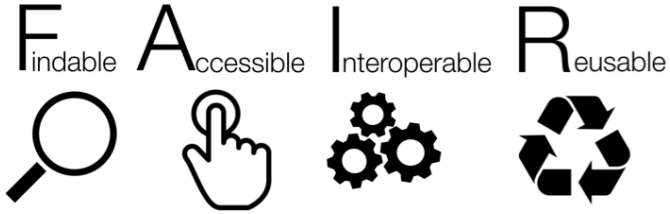
Tryggve activities

- Development of **sensitive data archiving technology**
- Development of **secure tools for analyzing sensitive data across borders**
- Operating a **use case program**
- Targeted development of the secure Tryggve platforms
- Implementing ELIXIR AAI based authentication and authorization solutions
- Providing assistance with GDPR related issues

Tryggve

neic

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 archiving activity in Tryggve

Tryggve

neic

Tryggve Sensitive data archiving activity

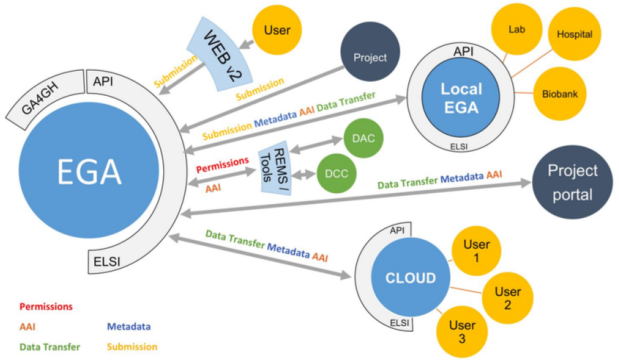
Develops technology and sets up pilots of federated sensitive data archiving, connected to European Genome-phenome Archive (EGA) infrastructure.

See presentation by Niclas Jareborg.

Federated EGA:

- Overcome legal obstacles for data deposition
  - Storage in the country of origin

- Federated querying across repositories
  - Increase the number of FAIR sensitive datas



Tryggve activities

- Development of **sensitive data archiving technology**
- Development of **secure tools for analyzing sensitive data across borders**
- Operating a **use case program**
- Targeted development of the secure Tryggve platforms
- Implementing ELIXIR AAI based authentication and authorization solutions
- Providing assistance with GDPR related issues

Tryggve

neic

Computing with distributed data

Emerging trend from the biomedical community use cases is how **to compute securely on distributed data sets**.

Tryggve currently working on three alternatives:

1. Pooling of data (technically simple but not always possible)
2. Federated model: compute close to data and combine the partial results (effectively separate processing)
3. Data streamed on-the-fly to joint analysis (data moves but is not stored outside the home repository)

Tryggve

neic

Tryggve activities

- Development of **sensitive data archiving technology**
- Development of **secure tools for analyzing sensitive data across borders**
- Operating a **use case program**
- Targeted development of the secure Tryggve platforms
- Implementing ELIXIR AAI based authentication and authorization solutions
- Providing assistance with GDPR related issues

Tryggve

neic

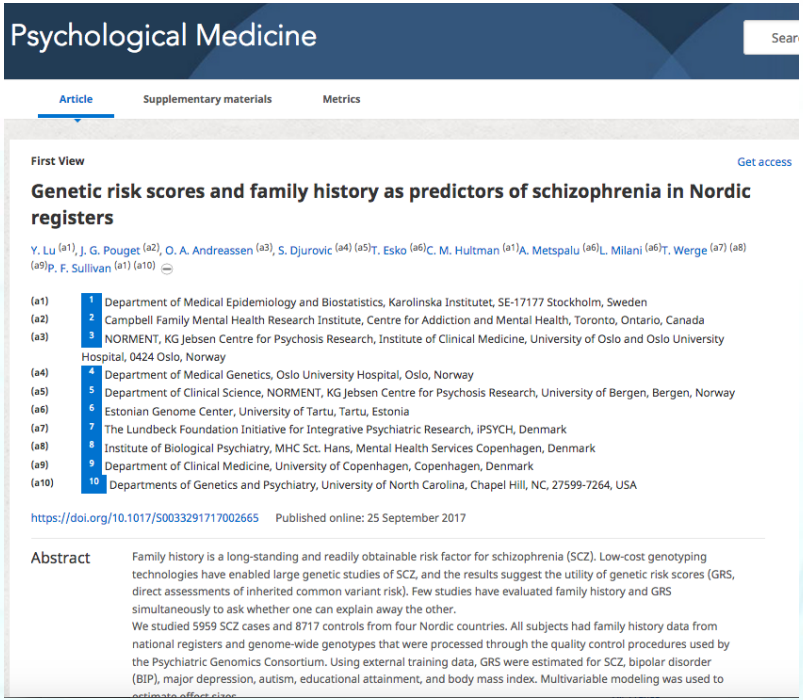
Tryggve use case program

- The Tryggve use cases are research or infrastructure service pilots that **drive the development** in the project with actual and concrete requirements.
- Use cases are a means to support researchers in taking the secure data services into use, and thereby to achieve results faster and safer.
- Use cases typically involve more than one Nordic country and involve cross-border collaboration on the focus areas of Tryggve.
- Use cases get expert support for solving their sensitive data challenges
    - Access to secure platforms is negotiated case by case

More information: neic.no/tryggve/usecase/

Tryggve

neic



Psychological Medicine

Article | Supplementary materials | Metrics

First View — Get access

Genetic risk scores and family history as predictors of schizophrenia in Nordic registers

Y. Lu [a1], J. G. Pouget [a2], O. A. Andreassen [a3], S. Djurovic [a4] [a5] T. Esko [a6] C. M. Hultman [a1] A. Metspalu [a6] L. Milani [a6] T. Werge [a7] [a8] [a9] P. F. Sullivan [a1] [a10]

[a1] 1 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-17177 Stockholm, Sweden
[a2] 2 Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada
[a3] 3 NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo and Oslo University Hospital, 0424 Oslo, Norway
[a4] 4 Department of Medical Genetics, Oslo University Hospital, Oslo, Norway
[a5] 5 Department of Clinical Science, NORMENT, KG Jebsen Centre for Psychosis Research, University of Bergen, Bergen, Norway
[a6] 6 Estonian Genome Center, University of Tartu, Tartu, Estonia
[a7] 7 The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark
[a8] 8 Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Denmark
[a9] 9 Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
[a10] 10 Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, 27599-7264, USA

https://doi.org/10.1017/S0033291717002665 Published online: 25 September 2017

Abstract

Family history is a long-standing and readily obtainable risk factor for schizophrenia (SCZ). Low-cost genotyping technologies have enabled large genetic studies of SCZ, and the results suggest the utility of genetic risk scores (GRS, direct assessments of inherited common variant risk). Few studies have evaluated family history and GRS simultaneously to ask whether one can explain away the other.

We studied 5959 SCZ cases and 8717 controls from four Nordic countries. All subjects had family history data from national registers and genome-wide genotypes that were processed through the quality control procedures used by the Psychiatric Genomics Consortium. Using external training data, GRS were estimated for SCZ, bipolar disorder (BIP), major depression, autism, educational attainment, and body mass index. Multivariable modeling was used to estimate effect sizes

Image sources:

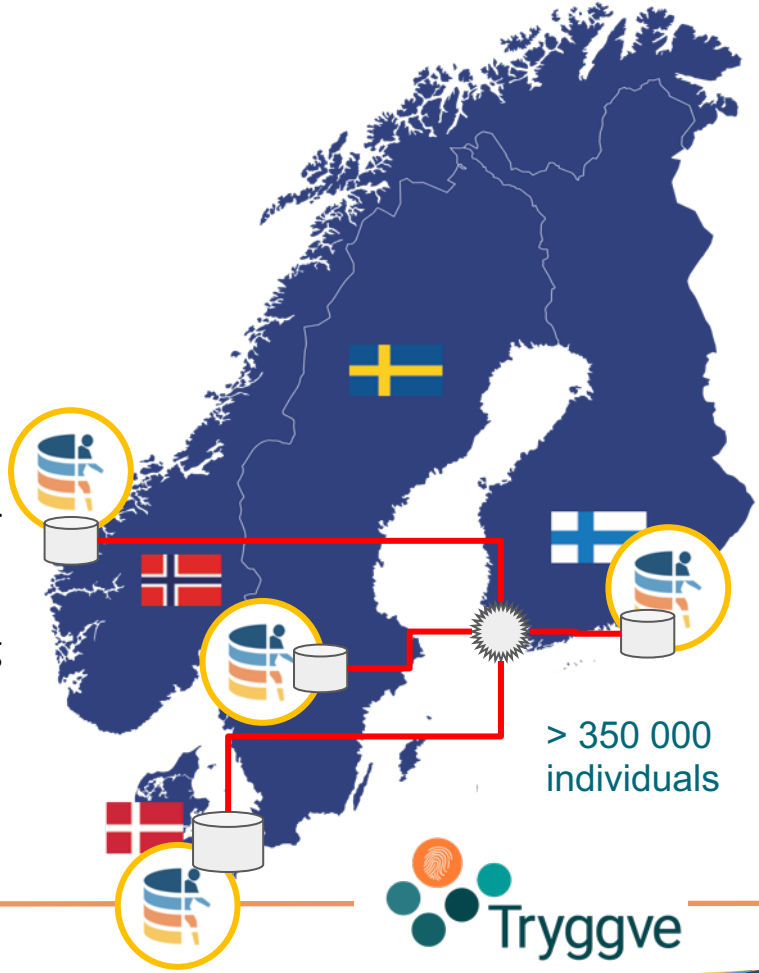
Processing sensitive data for schizophrenia research, In the field 04/2016, http://www.inthefieldstories.net/.

Genetics and environment across borders, NordForsk Magazine 2016.

Tryggve

Nordic Twin Study on Cancer

- Twin research on heritable and familial risk in prostate, breast, ovarian and colon cancers
- Cohort constructed by linking the population-based twin registries of Denmark, Finland, Norway and Sweden to their country-specific national cancer and cause-of-death registries. Genomic data also collected from the samples.
- A shared sensitive data processing environment required for method development and data harmonization
- Tryggve use case in progress presented by prof. J. Hjelmborg



NorTwinCan
Nordic Twin Study of Cancer

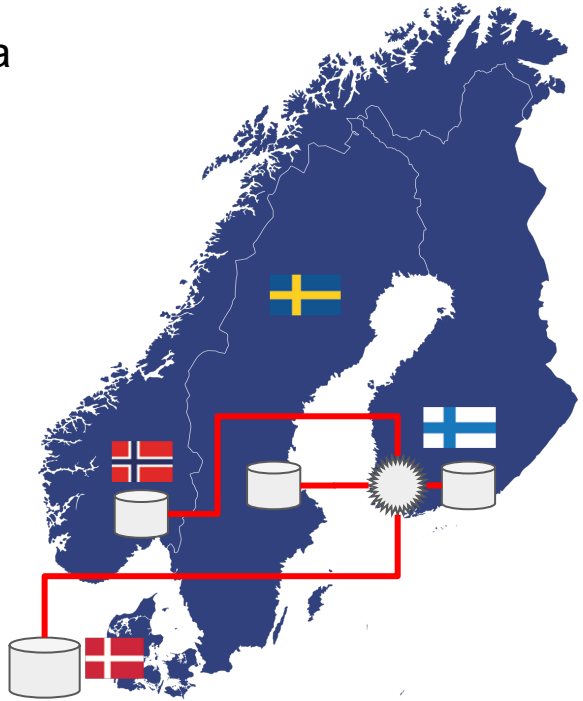
EUROPEAN GENOME-PHENOME ARCHIVE

http://nortwincan.org

Tryggve

Vision: Nordic federation of secure platforms

- Nordic infrastructure in which storage and computing form a **secure federated platform across borders**
    - Sensitive data stored in secure archives connected to the Federated EGA infrastructure
    - Distributed data made computationally available on any of the processing environments
    - Workloads distributed in containers with standard interfaces (e.g. GA4GH compatible cloud)
- Platform suitable **international large-scale data sharing initiatives**



Tryggve

neic

Thank you!

Project Manager:

Antti Pursula
antti.pursula@csc.fi

Scientific Manager:
Henric Zazzi
hzazzi@kth.se

Web: neic.no/tryggve
Subscribe to newsletter online

Twitter: @NeICTryggve

Local contact points:

Denmark: Ali Syed (DTU)
alisyed@cbs.dtu.dk

Finland: Antti Pursula (CSC)
antti.pursula@csc.fi

Norway: Abdulrahman Azab (UiO)
abdulrahman.azab@usit.uio.no

Sweden: Jonas Hagberg (NBIS)
jonas.hagberg@nbis.se

Contact: tryggve@neic.no