Tryggve -- Nordic Collaboration for Sensitive Data

Abdulrahman Azab
Senior Advisor, NeIC Executive Team
(Previously, Tryggve project manager)

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

Tryggve

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

- Tryggve 2014-2017: groundwork, building blocks, piloting
- Tryggve2 2017-2020: larger funding ~6M€, going towards service delivery, increased focus on use cases
- Tryggve3 (Heilsa, TryggveDottir) 2021-2024: Main focus on operational Federated EGA

https://neic.no/tryggve https://neic.no/heilsa/

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

Secure computing and data environments

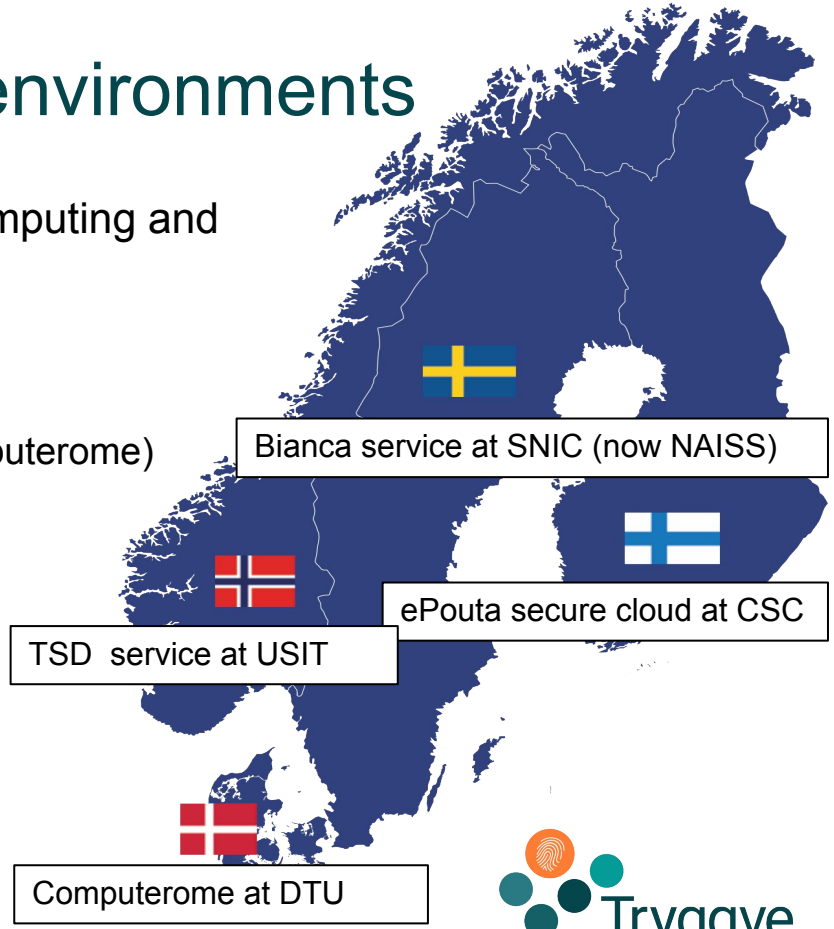
Tryggve service development relies on secure computing and data environments at participating ELIXIR Nodes

Secure cloud (Iaas)

- Infrastructure for data and computing (ePouta, Computerome)

Secure remote desktop (PaaS)

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



Building the cross-border sensitive data infrastructure

- Development of **sensitive data archiving technology**
- Development of **secure tools for analyzing sensitive data across borders**
- Operating a **use case program**
- Active **dissemination and outreach**

- Targeted development of the secure Tryggve platforms
- Implementing ELIXIR AAI based authentication and authorization solutions
- Providing assistance with GDPR related issues

Tryggve

neic

Development driven by research use cases

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

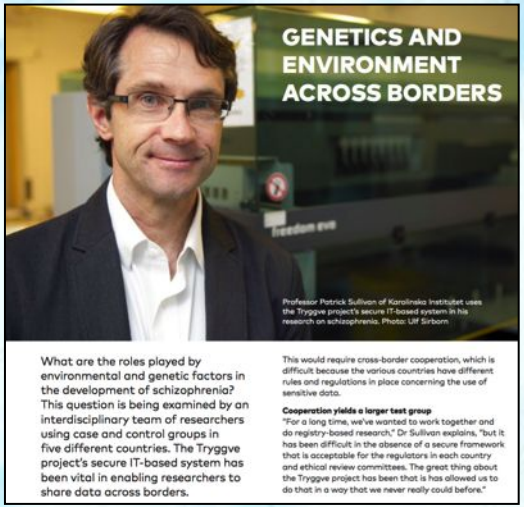
Family history is a long-standing and readily obtainable risk factor for schizophrenia (SCZ). Low-cost genotyping technologies have enabled large genetic studies of SCZ, and the results suggest the utility of genetic risk scores (GRS, direct assessments of inherited common variant risk). Few studies 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

Image sources:

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

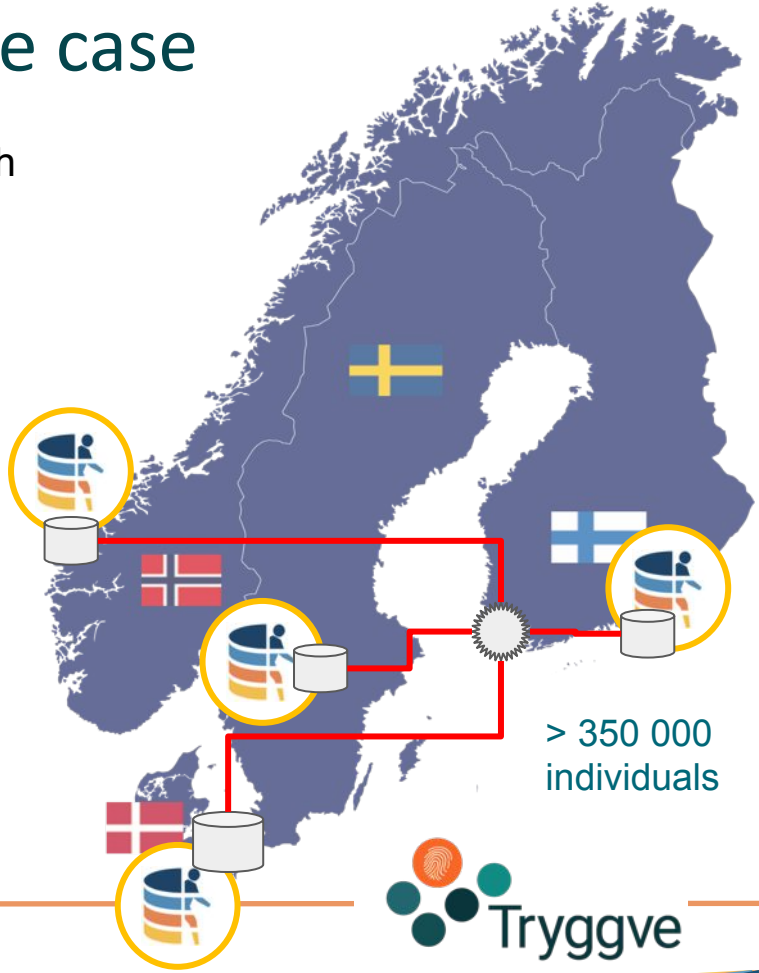
Genetics and environment across borders, NordForsk Magazine 2016.



Tryggve

Nordic Twin Study on Cancer - Use case

- Largest twin study in the world available for the research on heritable and familial risk of 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.
- Shared sensitive data processing environment required for method development and novel analyses
- Tryggve use case



NorTwinCan
Nordic Twin Study of Cancer

http://nortwincan.org

EUROPEAN GENOME-PHENOME ARCHIVE

Key activities and results - I

Secure and FAIR storage of sensitive data in the Nordics

- Purpose:
  - Enable secure repositories for sensitive data in each participating country interoperable with the European Genome-phenome Archive (EGA) infrastructure
- Key results:
  - Produced software for the secure storage service and deployed the first-in-Europe Federated EGA node prototype in 3 countries (FI, NO, SE)
- Next steps:
  - Moving from prototype stage to operations stage
  - Developing standalone repository for cases where connection to EGA is not required

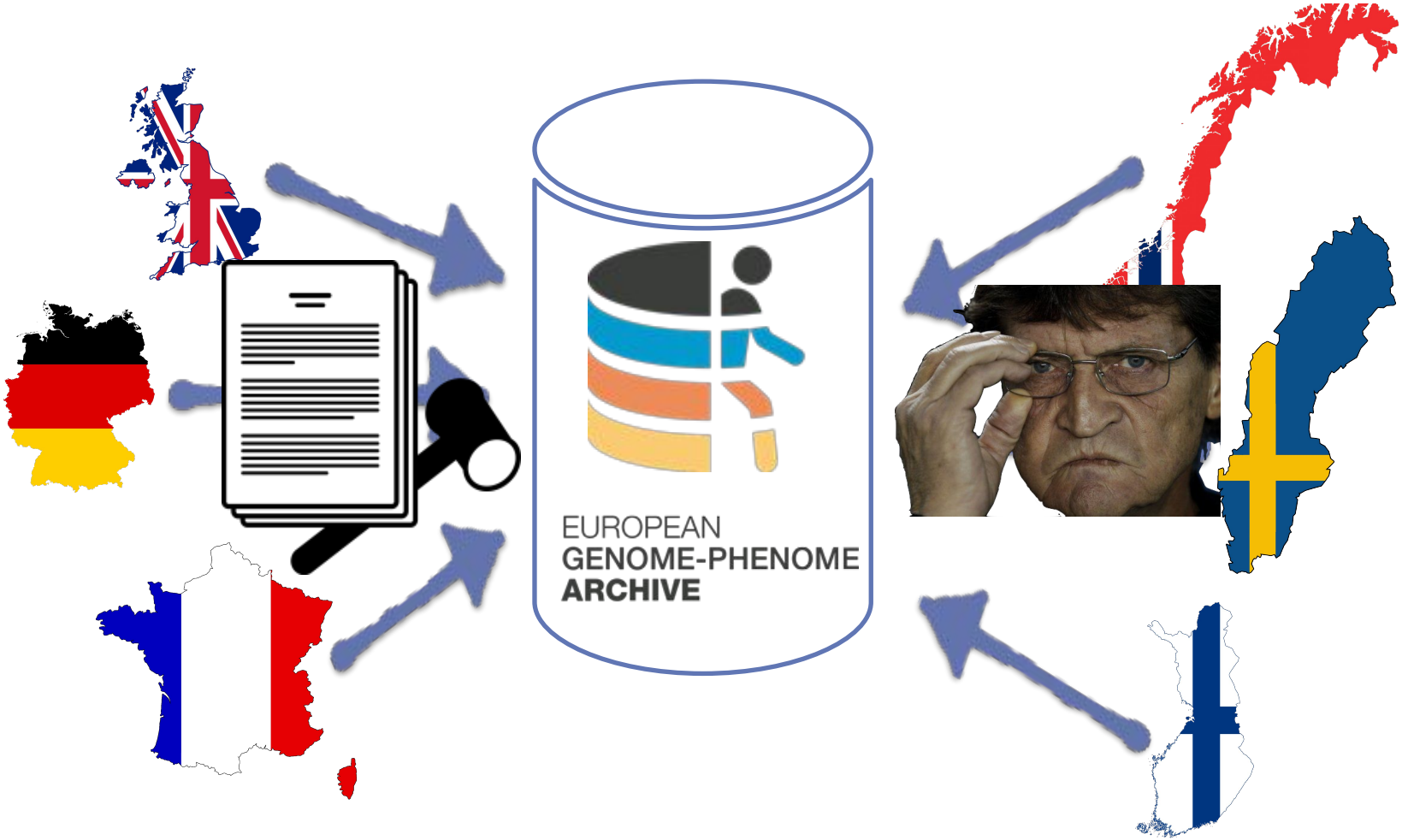
neic Tryggve

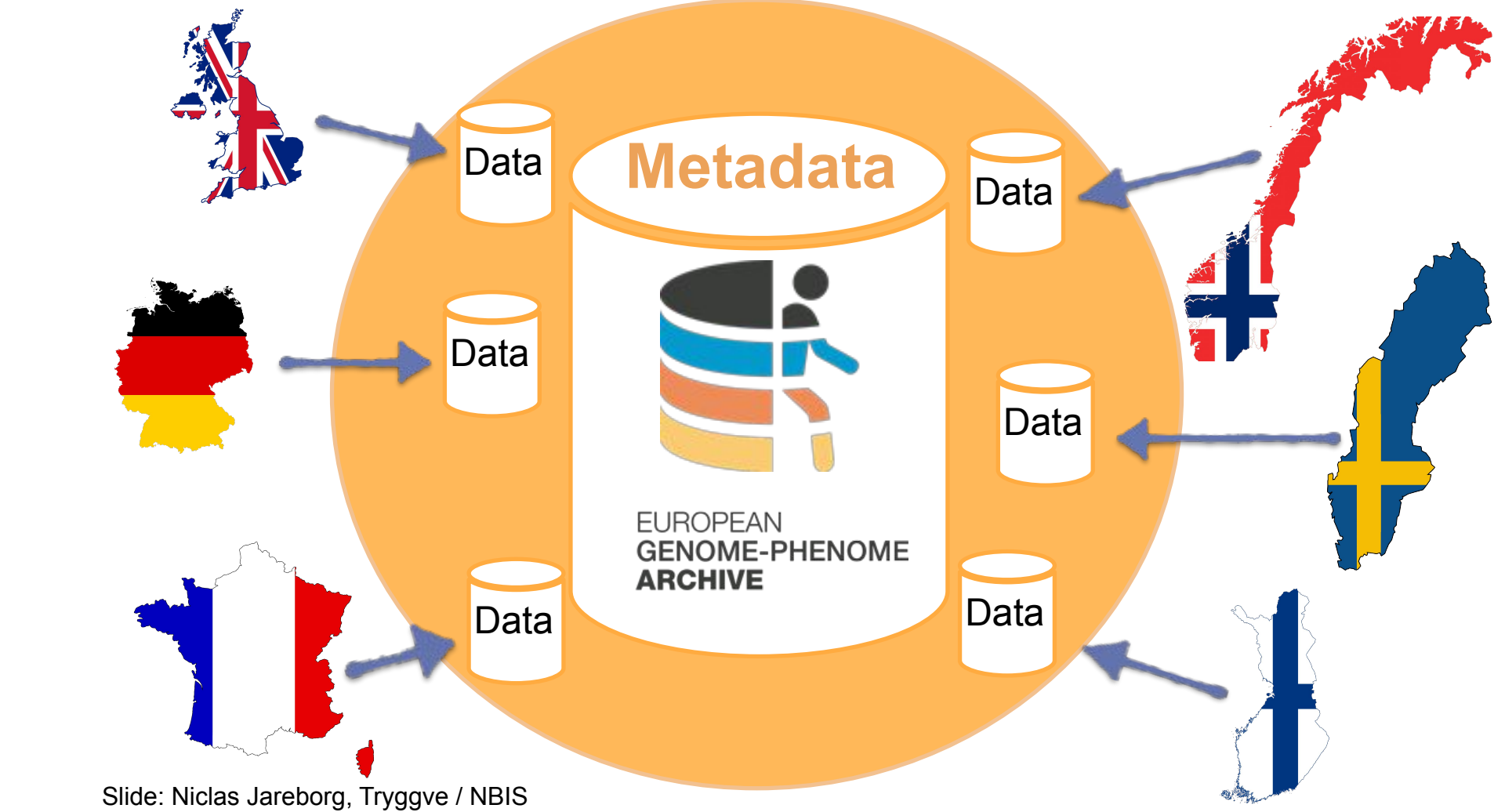
European Genome-phenome Archive EGA

- Permanent secure archiving ar sharing of all types of potentially identifiable human genetic and phenotypic data

- 'FAIR' (Findable, Accessible, Interoperable, and Reusable) principles

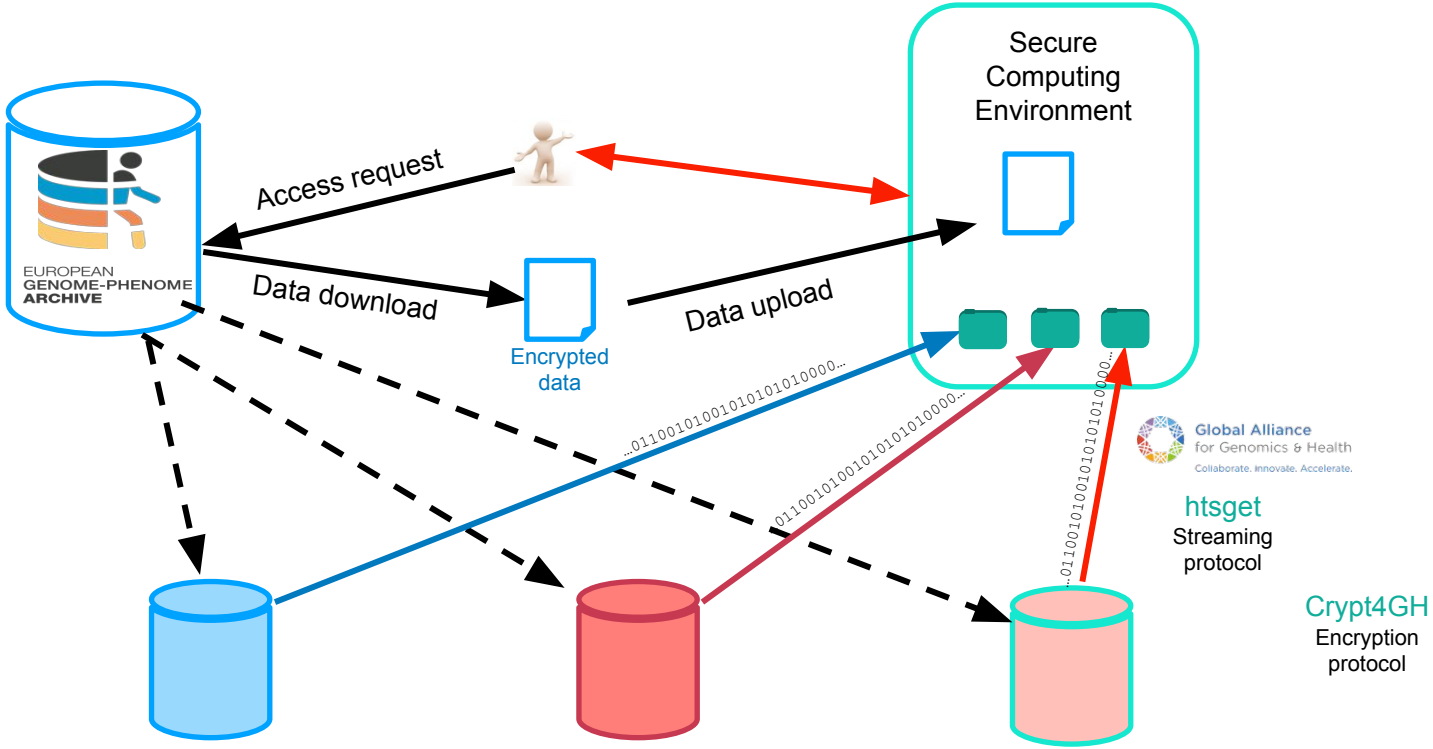
Slide: Niclas Jareborg, Tryggve / NBIS


EUROPEAN
GENOME-PHENOME
ARCHIVE



Slide: Niclas Jareborg, Tryggve / NBIS

Future ways of accessing data



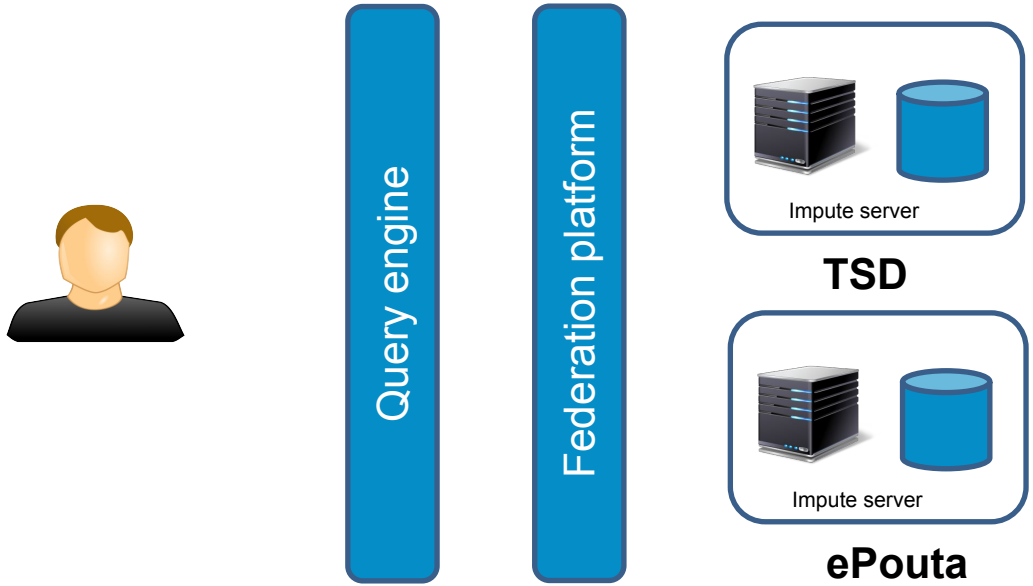
Key activities and results - II

Cross-border workflows for processing Nordic sensitive data

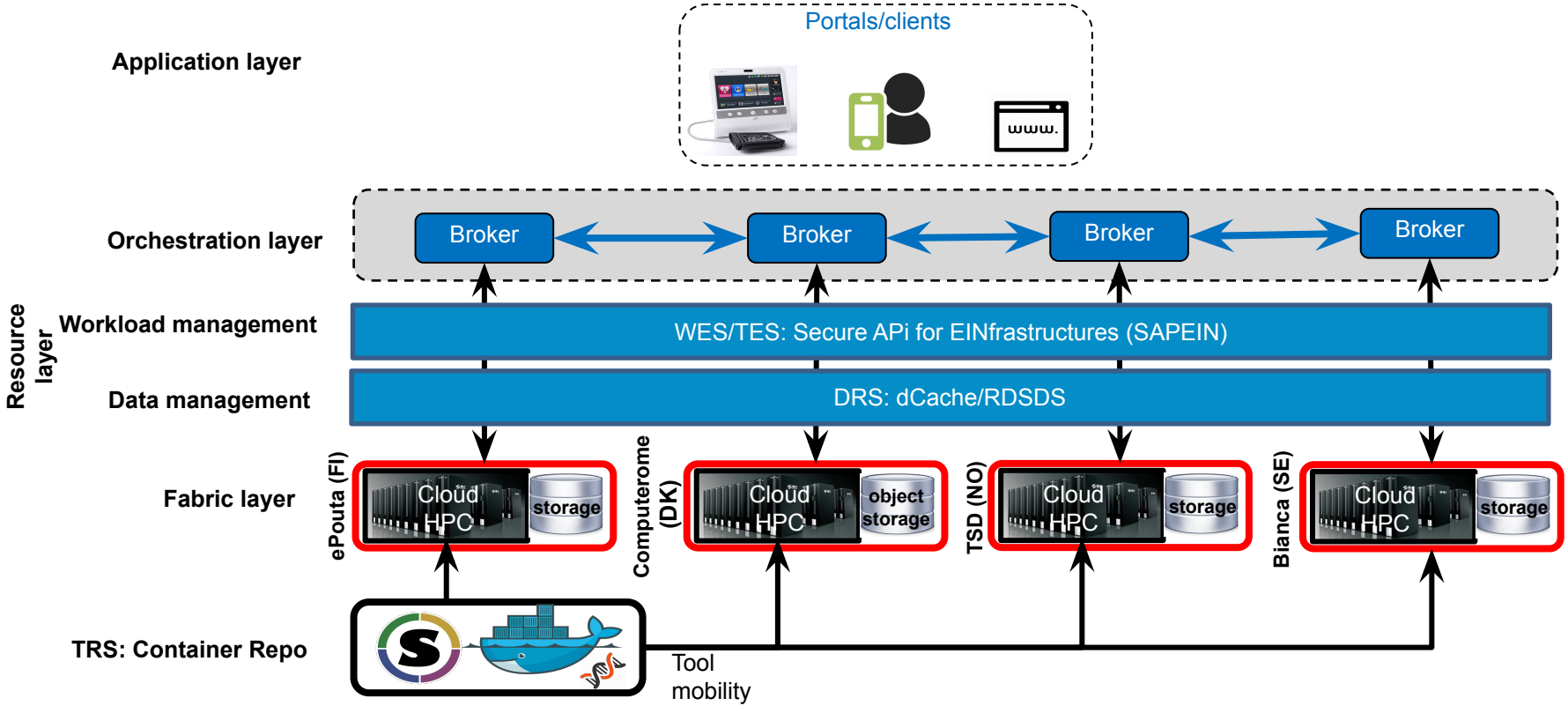
- Purpose:
  - Enable use cases that involve analysis of sensitive data that is stored in different locations and countries.
    - The main usage scenarios are the **Joint processing** (“bring data to compute”) and the **Federated processing** (“bring compute to data”) approaches.
- Key results:
  - Demonstrated the on demand access method for data stored in FI and SE.
  - Studied and described secure cloud and workflow technologies.

neic Tryggve

Use case 2-9 Impute server – Distributed imputation



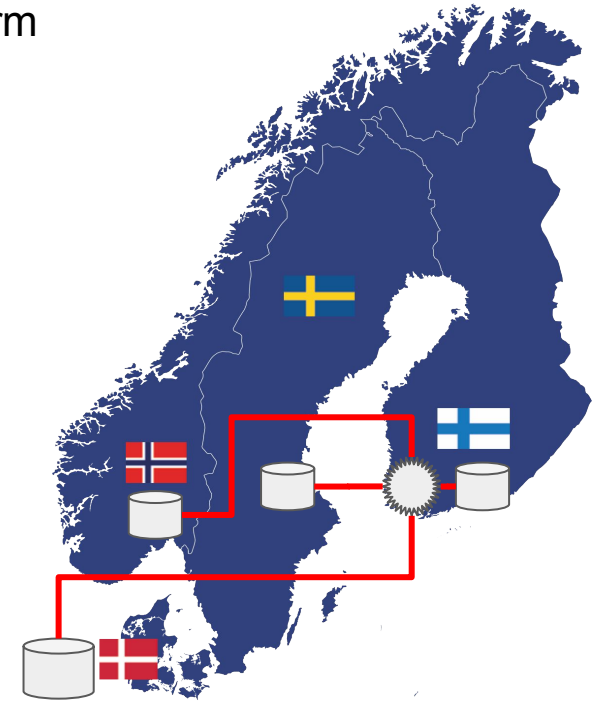
Nordic sensitive data e-Infrastructure federation – GA4GH compliant



Vision for cross-border sensitive data infrastructure

Aiming at Nordic infrastructure in which storage & computing form a **secure distributed platform** for sensitive data:

- The user can **access distributed data from any node** of the platform.
- **Workloads are distributed in containers** with standard GA4GH workflow management and interfaces.
- Sensitive data can be **stored in secure repositories** connected to the European Federated EGA infrastructure

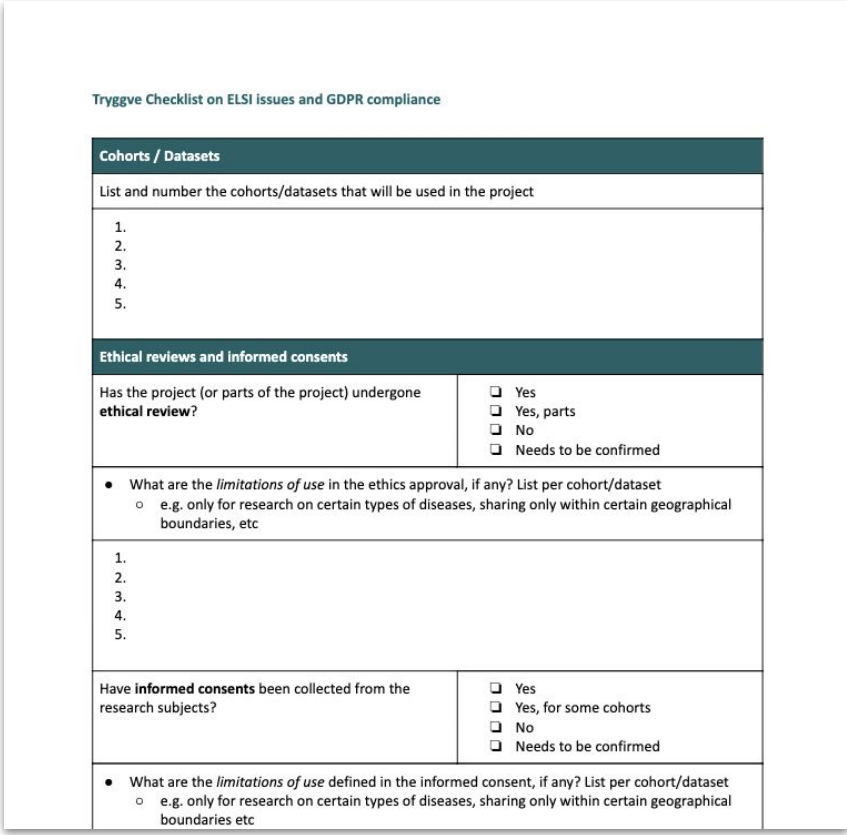


Tryggve

neic

Tryggve ELSI “Checklist”

- ELSI issues & GDPR compliance
  - Ethical reviews & Informed consents
  - Controllers
  - Legal basis
  - Data processing agreements
  - Data Protection Impact Assessments
  - Data sharing
  - Other legal considerations

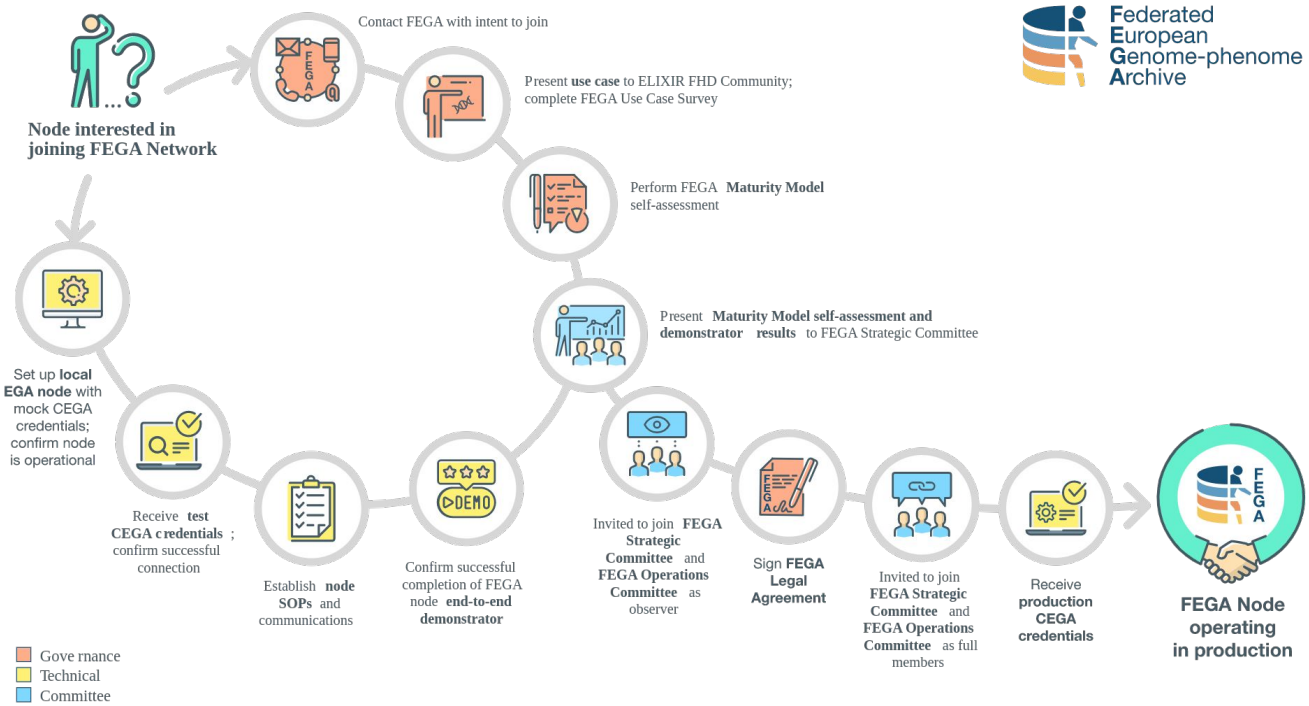


Tryggve Checklist on ELSI issues and GDPR compliance

Cohorts / Datasets	
List and number the cohorts/datasets that will be used in the project	
1. 2. 3. 4. 5.	
Ethical reviews and informed consents	
Has the project (or parts of the project) undergone **ethical review**?	❑ Yes ❑ Yes, parts ❑ No ❑ Needs to be confirmed
• What are the *limitations of use* in the ethics approval, if any? List per cohort/dataset ○ e.g. only for research on certain types of diseases, sharing only within certain geographical boundaries, etc	
1. 2. 3. 4. 5.	
Have **informed consents** been collected from the research subjects?	❑ Yes ❑ Yes, for some cohorts ❑ No ❑ Needs to be confirmed
• What are the *limitations of use* defined in the informed consent, if any? List per cohort/dataset ○ e.g. only for research on certain types of diseases, sharing only within certain geographical boundaries etc	

https://docs.google.com/document/d/1Kg_IrKc0KJVCvMw4iyEhrRsIlenEjbjKkWbZ1aK5J68/edit

Federated EGA - Journey*



Federated European Genome-phenome Archive

* as illustrated by https://ega-archive.github.io/FEGA-onboarding

NeIC Heilsa Tryggvedottir - joining Federated EGA

- A code base for archiving of sensitive data
  - Documentation available at: https://neic-sda.readthedocs.io/en/latest/
  - Packaged to be convenient for setting up Federated EGA, or stand-alone instances for archiving of sensitive data
  - Technical questions about the solution https://gitter.im/neicnordic/sda-pipeline
- Common operating procedures for (Federated EGA) nodes
  - Documentation
  - Processes
  - Roles and responsibilities
  - Security

Nordic-Baltic Sensitive Data forum

- Nordic-Baltic Sensitive Data Forum brings together researchers, **service providers**, **security experts**, **service developers/administrators**, **data protection experts**, **lawyers**, **policymakers**, and **activity funders** in one community to learn from each other, plan for competence-building (e.g. training events and workshops), and collaboration activities (including Nordic/EU-funded projects).

- Subscribe: https://neic.no/mailman/listinfo/sensitive-data
- Contact: sensitive-data@neic.no

neic Tryggve

Thank you!

Project Manager:

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

abdulrahman.azab@sigma2.no (from July)

Web:
neic.no/tryggve
neic.no/heilsa
sensitive-data@neic.no