Nordic collaboration for sensitive data
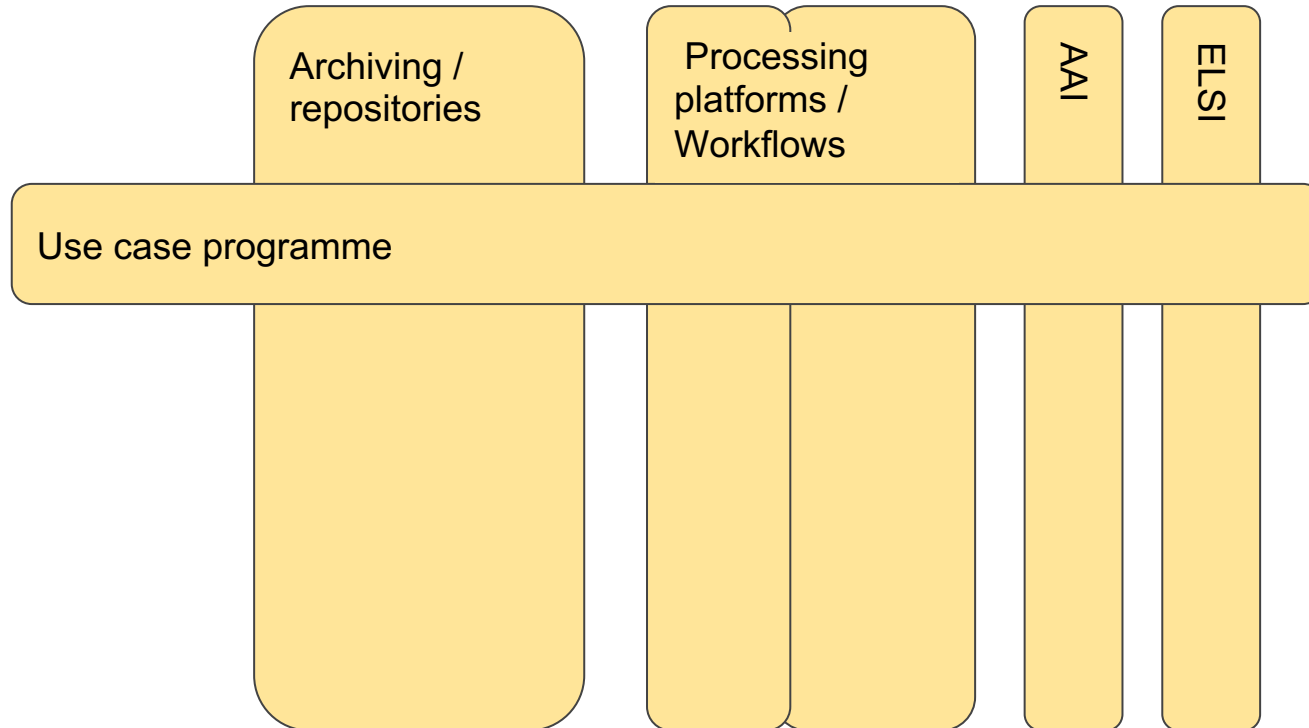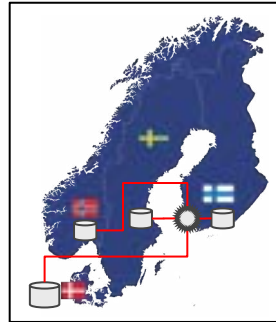
Abdulrahman Azab

Some slides from Antti pursula

NeIC projects scope

Archiving / repositories

Processing platforms / Workflows

AAI

ELSI

Use case programme

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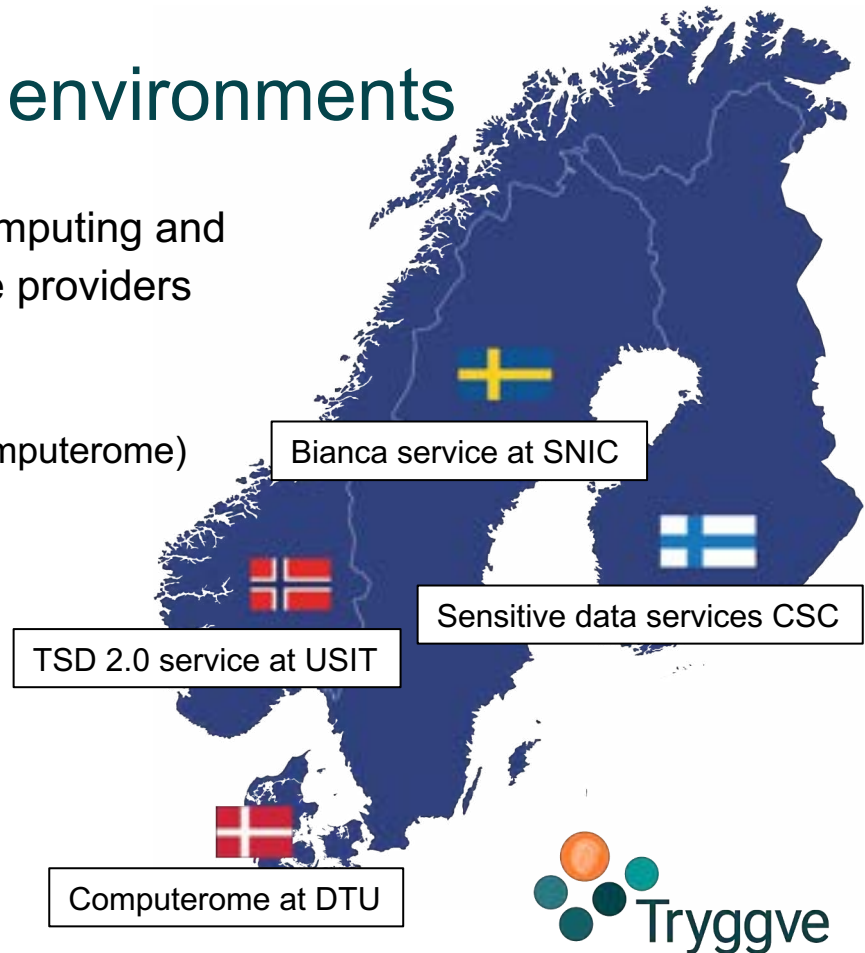

Tryggve

Secure computing and data environments

Nordic service development relies on secure computing and data environments by the nordic e-Infrastructure providers
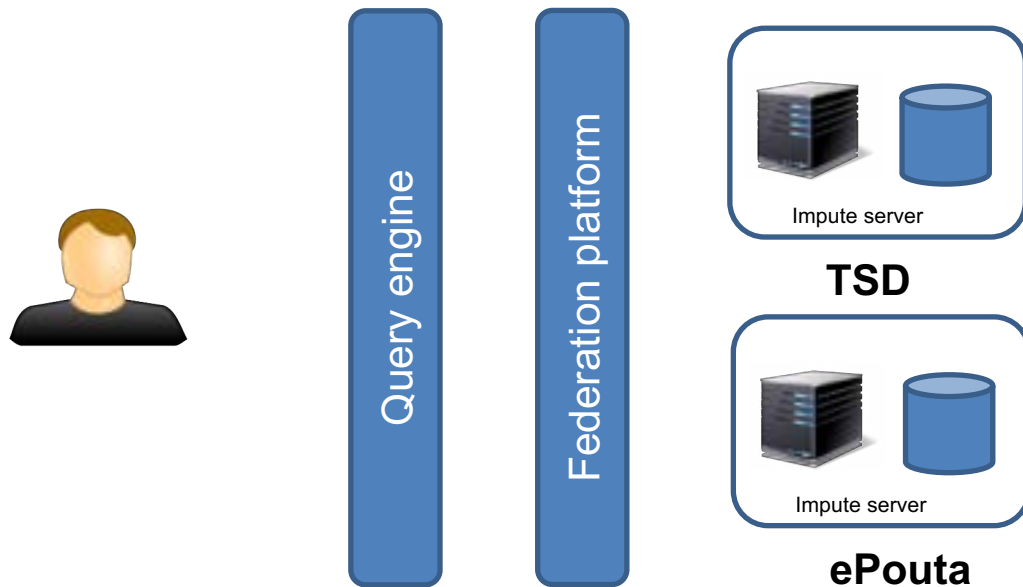
Secure cloud (IaaS)
- Infrastructure for data and computing (ePouta, Computerome)

Secure remote desktop (PaaS)
- Backed up with computing resources (TSD, Computerome, Bianca, CSC ePouta)

Bianca service at SNIC

Sensitive data services CSC

TSD 2.0 service at USIT

Computerome at DTU

Tryggve

Use case 2-9 Impute server – Distributed imputation

Tryggve ELSI aspects of current use cases

- Based on the analysis of Tryggve use cases certain commonalities regarding ELSI topics can be found, mainly the following:
  - Possible specific requirements of consents towards data processing.
  - Determining the data controller
  - Uncertainties regarding the legal basis for data processing.
  - Data processor contracts with sub-processors.
  - Influence of local non-GDPR legislation.
  - Terms for accessing and moving register data (esp. from Denmark).
- Summary document
- *Fed into work on a checklist*

Tryggve Training event for Nordic researchers on ELSI topics

"Tryggve workshop on Implications of GDPR in cross-border research in the Nordics"

- Helsinki, April 2019

- Applying the GDPR in Nordic research projects

- BBMRI Guidelines and Code of Conduct for health research

- GDPR issues for cross-border projects - Checklist
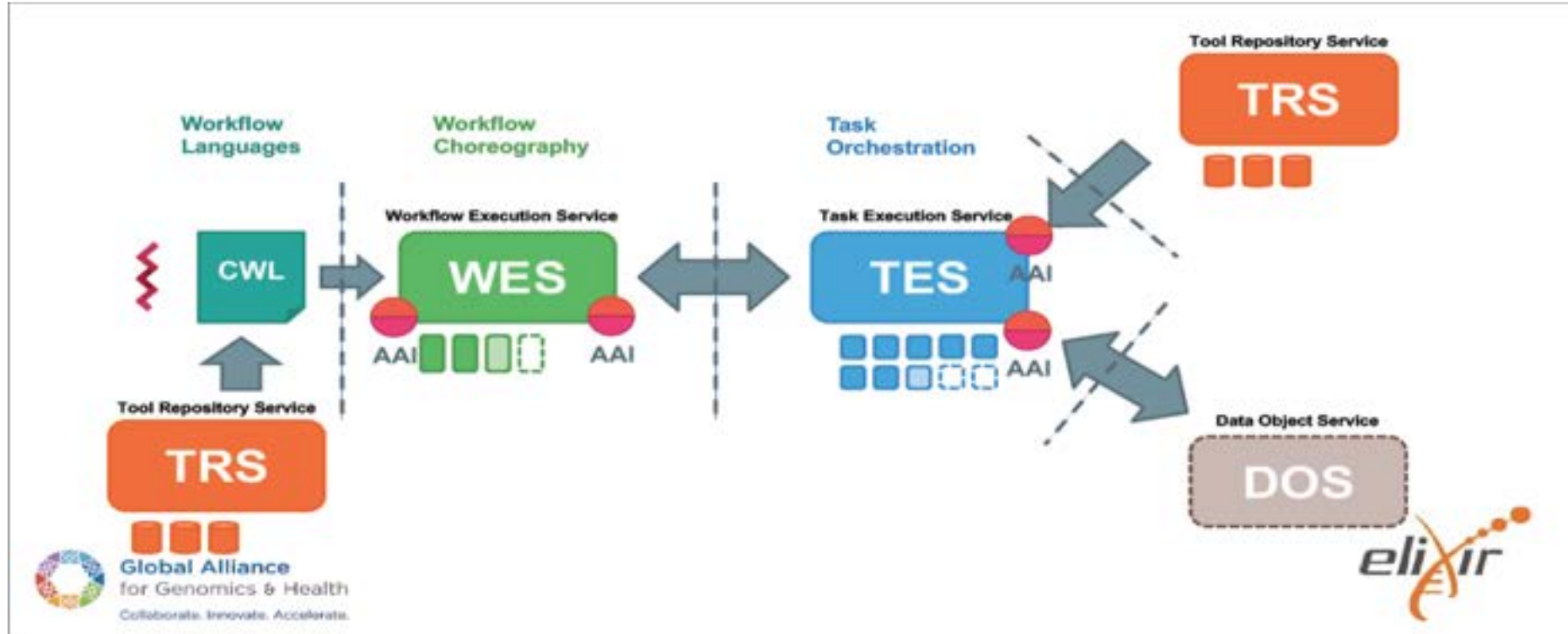
Tryggve ELSI "Checklist"

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

https://docs.google.com/document/d/1Kg_lrKc0KJVCvMw4iyEhrRslIenEjbjKkWbZ1aK5J68/edit

GA4GH Compatible Platform (Functional Architecture)
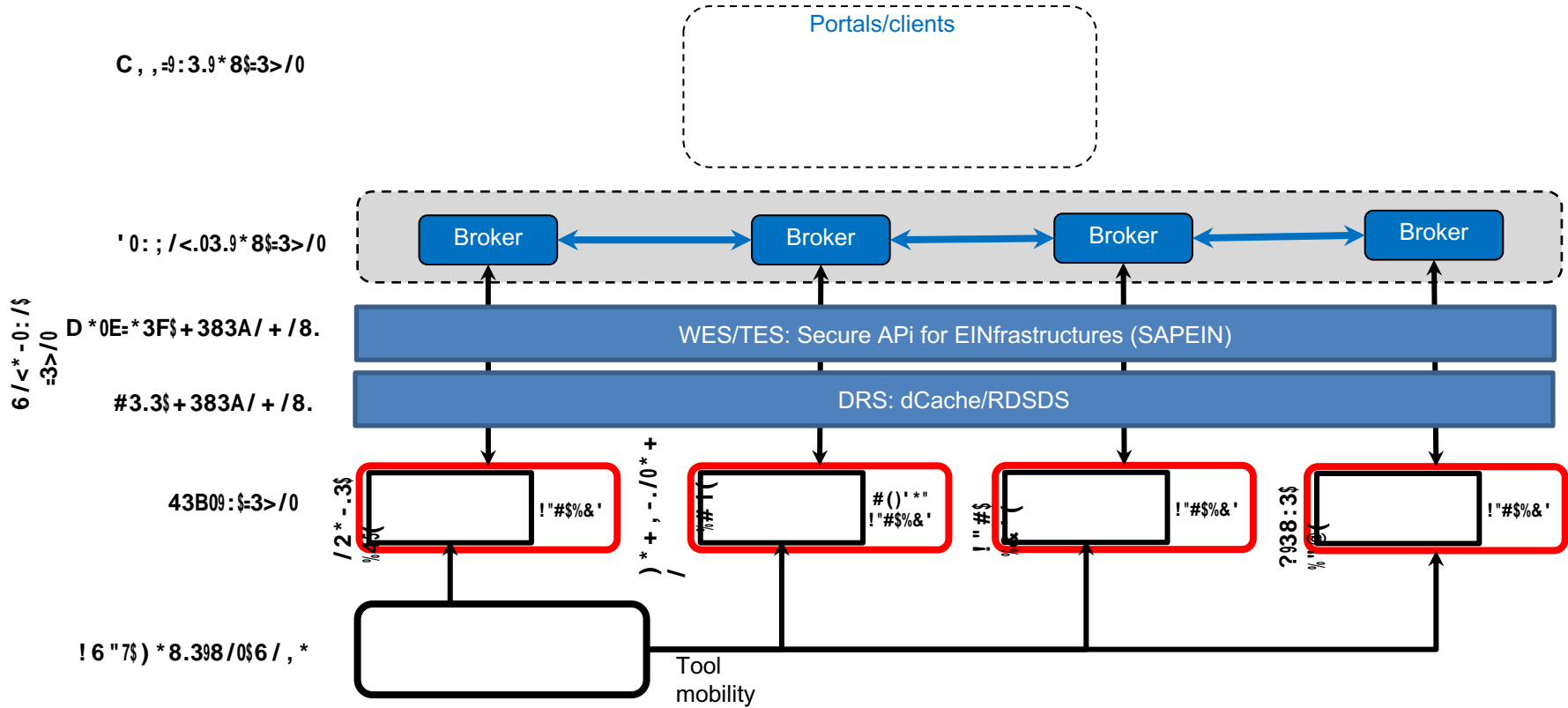
Are we there?

- No we are not!

- What do we need:

  - **Service providers to trust each other:** Security, Data protection, Level of assurance

  - **Data owners/providers** to trust and collaborate with infrastructure providers

  - **ELSI Checklist/ best practices** for cross-border use cases

◈neic

Nordic-Baltic Forum for Building the cross-border sensitive data infrastructure

- Development of **sensitive data sharing and archiving technology**

- Support the Development of secure workflows for **processing sensitive data across borders**

- Define cross-border **use cases to support**

- Further development of the **secure Nordic platforms**

- Implementing LifeScience AAI based authentication and authorization solutions

- Providing assistance with GDPR/legistlation related issues

◈ neic

Draft Call: HORIZON-INFRA-2023-EOSC-01-06: Trusted environments for sensitive data management in EOSC (RIA, 15M€, 3p, 100%, 22/12/2022-28/03/2023)

Expected Outcome: Project results are expected to contribute to all the following expected outcomes:

- **Expansion of EOSC's access to resources provided by public authorities**, including national agencies, in highly sensitive areas, such as the health sector, governmental statistics or geo-spatial applications, that ensures the opening of these valuable data sets for novel research through a standard set of methods able to effectively enable sensitive data sharing/processing/analysis;

- Emergence of trusted environments for **management and sharing of sensitive data** in order to facilitate new ways of using sensitive data sets;

- Demonstration that **FAIR data workflows with sensitive** data are securely possible and can benefit both the data providers and the wider science community.

Upcoming calls

Exploring the Innovation Potential In Patient Journals

This Autumn, Nordic Innovation will launch a Call For Proposals for a project which seeks to showcase the innovation potential in Nordic patient journals.

Published: 20.09.2022 Updated: 21.09.2022

**The call will be open within the next six weeks.**

The aim of this call is to fund a project demonstrating a concrete solution utilizing the patient journals across Nordic countries in solving healthcare challenges.

Potential funding:	Up to 6.000.000 NOK

In addition to the potential funding, there will be a co-financing requirement for the winning consortia.

Upcoming calls
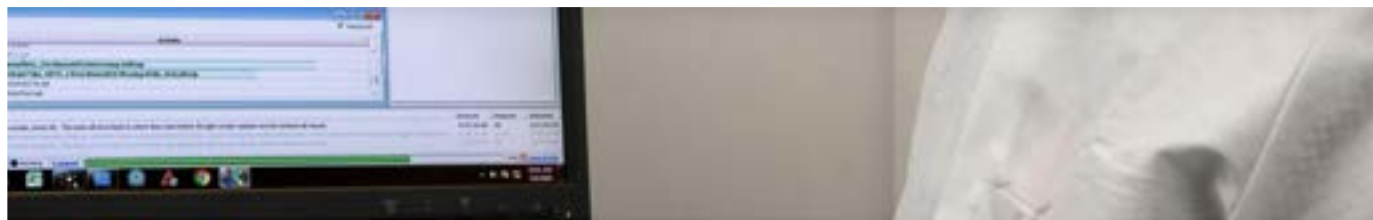
Developing an AI-solution For Nordic Patient Journals

This Autumn, Nordic Innovation will look for a consortia of partners who can develop an AI-solution capable of processing data from patient summaries across the Nordics.

Published: 20.09.2022 Updated: 20.09.2022

The call will be open within the next six weeks.

Potential funding:	Up to 6.000.000 NOK

In addition to the potential funding, there will be a co-financing requirement for the winning consortia.

Nordic-Baltic Forum for Building the cross-border sensitive data infrastructure - NeIC

- Community effort

- Some overhead funding (F2F meetings, etc)

- Analyses for future collaboration opportunities

- Maintaining collaboration platforms through small funded projects and activities

- Establishment and maintenance of collaboration networks (slack, mailing list)

- Ideas?

neic

Yet another brainstorming on sensitive data sharing!

Kahoot!
1078418

Cross boarder sensitive data sharing: what are the bottlenecks?

- World Cloud

- Discussion on the outcome

What are the most suitable directions of intervention to overcome the bottlenecks?

- Multiple choices:
  1. Coordinate investments toward the establishment of a suitable (cross boarders dimensioned) data infrastructure?
  2. Enhance coordination toward the establishment of a common legal framework
  3. Facilitate dialogue between sensitive data service providers – data owners – data consumers to resolve case by case studies
  4. None of these

What are the most viable directions of intervention to overcome the bottlenecks?

- Multiple choices:
    1. Coordinate investments toward the establishment of a suitable (cross boarders dimensioned) data infrastructure?
    2. Enhance coordination toward the establishment of a common legal framework
    3. Facilitate dialogue between sensitive data service providers – data owners – data consumers to resolve case by case studies
    4. None of these

Bottom up or top down?

1. It is better to start from a top down (case by case approach)
2. or a bottom up?

From the data owner perspective:
What is needed for a data owner to a certain country to allow data sharing and/ or data processing in an other country?

1. It depends on the type of data processing/sharing. Es.: streaming, imputing, giving away data pose different legal and technological requirements

2. It is only about having the proper legal framework, regardless the technology. The receiving hand will take care of it!

3. It is only about having trust-worthy technological solutions: If I am the data owner I want to know who is the technology provider at the receiving hand, no matter what the law says!
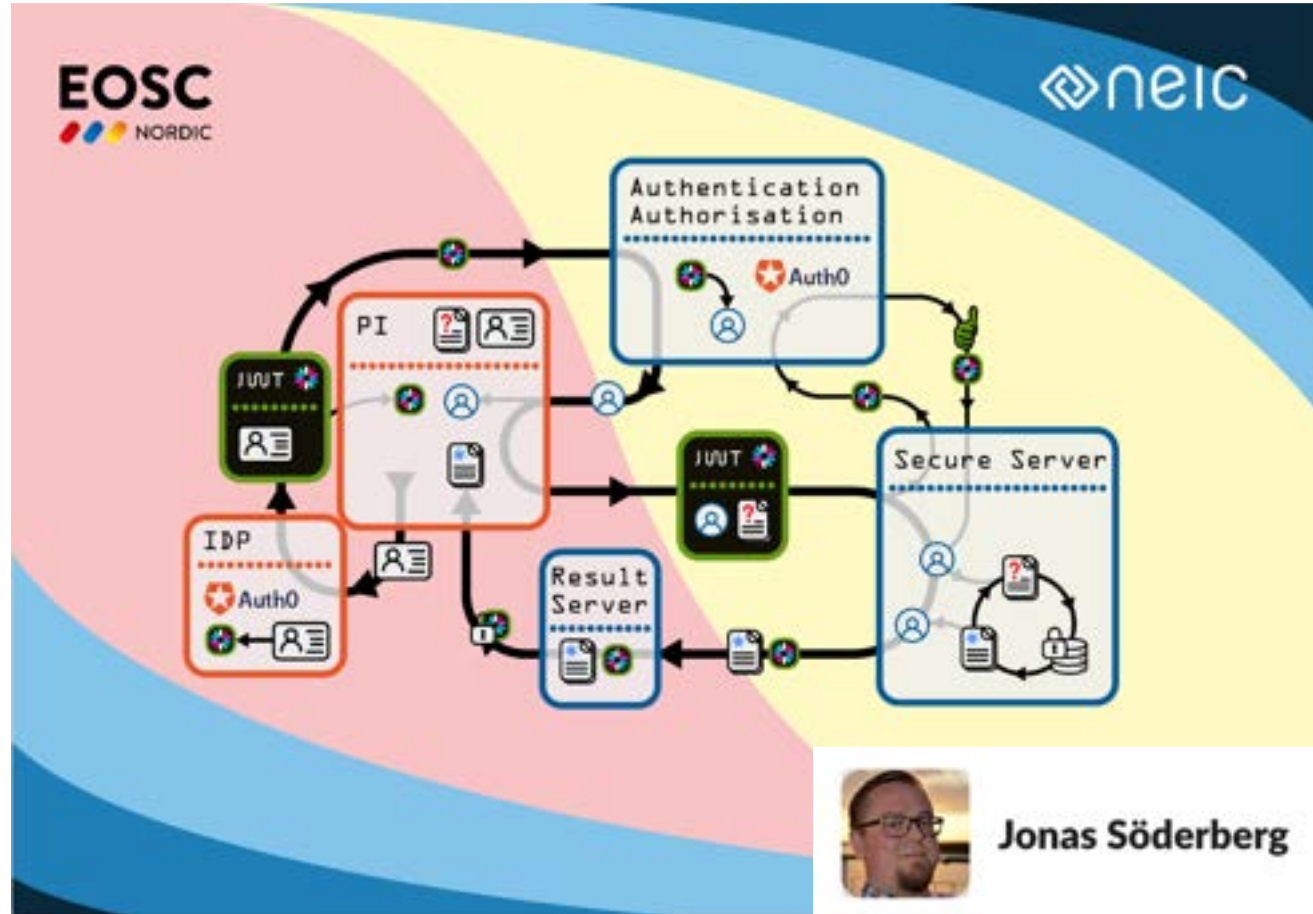
How to promote trust?

- Brainstorming – World Cloud

How to promote the Nordic trust?

- Brainstorm

Support slides

EOSC-Nordic's Technical vision for federated data access

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 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

  https://neic.no/tryggve

D2/D2 Secure platforms and workflows across

- 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.

- Proof of concept implementations with use cases
- Follow global standards from GA4GH

Use case programme

- Direct benefit for Nordic biomedical and health research
  - Use cases function as drivers for development of sensitive data infrastructure
- 12 use cases in progress

- Next steps:
  - Produce concrete results in use cases together with the users

Summary of use cases

Henric Zazzi
Tryggve Scientific Manager

Criteria for use cases

- Use cases are a **key component** in the Tryggve2 project
  - Demonstrate benefit for users
  - **Drive development**
  - Effective dissemination for the project

- Process for managing use cases should account for
  - Fair evaluation of cases
  - Alignment with strategy of partners
  - Decision-making roles
  - Transparency within and outside project

Managing use cases
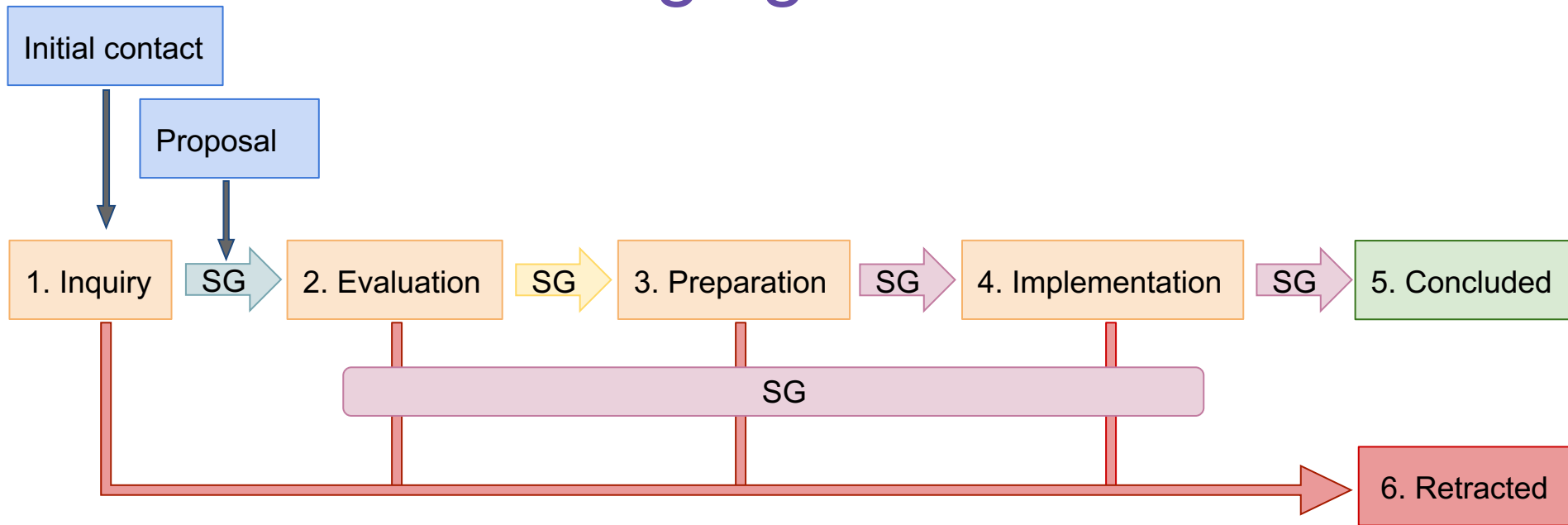
Labels signaling concerns

| **Pending**	Minor concerns
On Hold	Concerns, potentially damaging
Stalled	Serious concerns, threatening overall performance. Could be **retracted** if agreed upon by *Customer PI*

Tryggve ◈neic

High priority labels

- use case that could significantly benefit tryggve

- Scientific Manager should be present on project meetings

- Prioritized on use case management meetings

Criteria
Novel type of use case
Drives new technical innovation
High impact in the scientific community

Use case status

Number of use cases
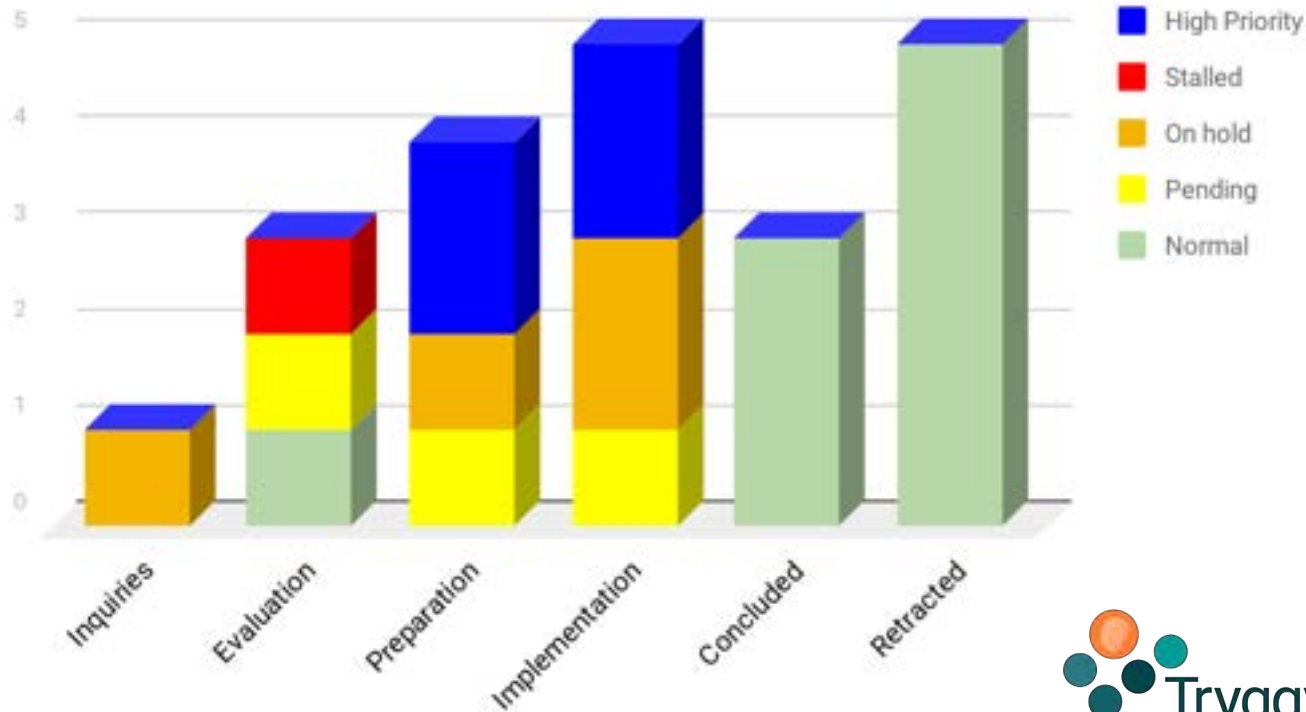
quiries

+ Add a card

Evaluation ⋯

Pending
1-6 STAGING
👁 🗐 💬 4   A 👤

Stalled
2-8 Genetics
🗐   👤

+ Add another card

Preparation ⋯

1-10 NordTreat
🗐   👤

Pending
2-7 Consumer data
🗐 💬 4   A 👤

+ Add another card

Implementation ⋯

High priority
1-1 Schizophrenia
👁 🗐 💬 3 🗂 2/3
👤👤👤👤👤

Pending
1-3 I-SCAN
👁   👤👤👤

High priority
2-9 Impute
🗐 🗂 2/4   👤 A 👤

On hold
2-1 Databank
👁 💬 6   A 👤

Stalled
2-5 PhenoMeNal
👁 🗐 💬 6   👤👤

High priority
1-4 NorTwinCan
👁 🗐 💬 8   👤 👤 MA

2-2 CRC
👁 🗐 💬 3   👤

High priority
1-9 Neoepitope pipeline

+ Add another card

Concluded ⋯

Tryggve1-1 Sullivan
🗐

Tryggve1-2 Mälarstig
🗐 💬 1

Tryggve1-3 Dillner
🗐

+ Add another card

Retracted

1-5 ScopeOMT
🗐 💬 3

2-6 Multiresistan

2-3 Data transfer
🗐

2-4 Cloud extens
🗐

1-8 SwEstGen
👁 🗐 💬 3

1-7 SafeTwoTreat
👁 🗐 💬 1

+ Add another c

Tryggve technical solutions

- Access to national sensitive data infrastructure
  - Remote desktop for access of data

- Storing sensitive data

- Sharing sensitive data
  - Secure data transfer
  - Legal aspect for transferring data across nordic countries

- Analyze sensitive data in a secure environment

Development driven by research use cases

Processing sensitive data for schizophrenia research

GENETICS AND ENVIRONMENT ACROSS BORDERS

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.

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 in progress
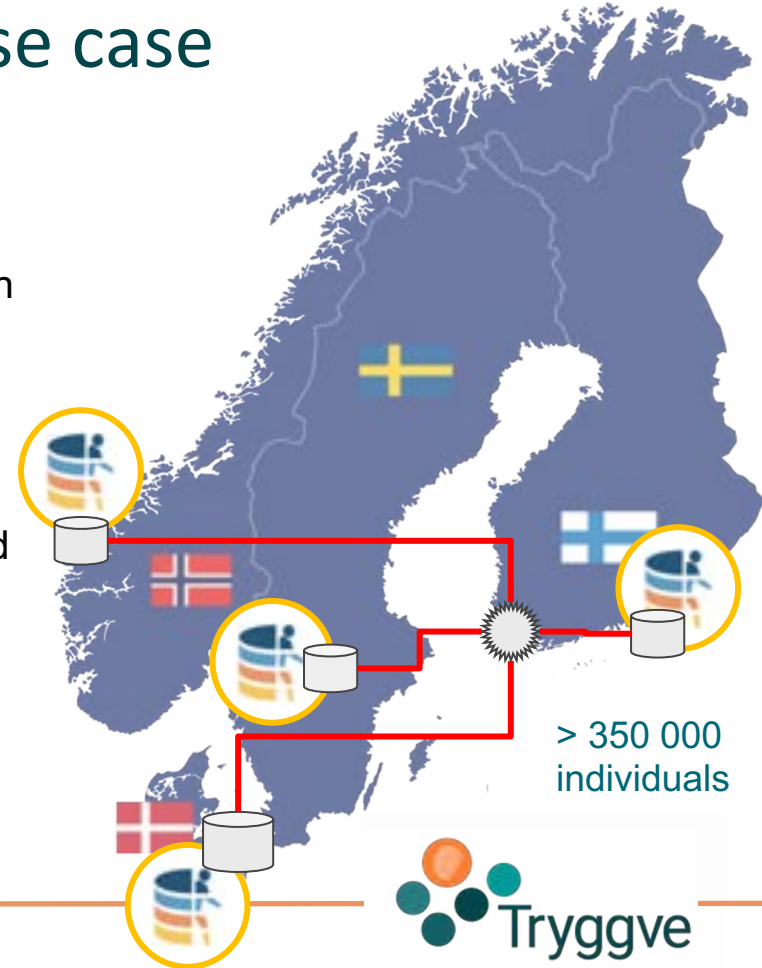
> 350 000 individuals

NorTwinCan
Nordic Twin Study of Cancer

http://nortwincan.org

EUROPEAN GENOME-PHENOME ARCHIVE

Tryggve

Building the cross-border sensitive data infrastructure

- Development of **sensitive data archiving technology**

- Development of secure workflows for **processing 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

D8 Outreach and dissemination

Raise awareness within stakeholders on the outcomes of Tryggve

- Tryggve monthly Newsletter
  - Anyone can suggest and/or write brief updates
- Twitter @NeICTryggve

- Presentations and posters

D8 Outreach

NeIC
@NeICnordic

The I-SCAN study explores the risk of cancer in pa
with inflammatory bowel disease. With the help of
@NeICTryggve, the study takes advantage of Nor
health registries and aims to provide better estima
cancer risk in IBD. Read our article: neic.no/news/
/01/2...

neic

2020-01-28

THE I-SCAN STUDY AS A TRYGGVE USE CASE

I-SCAN is a study exploring the risk of cancer in patients with inflammatory bowel disease. Generally, there has been a belief that patients with IBD have an increased risk of cancer, but studies have shown that for many patients it may not be increased at all. The collaborative project, with the help of Tryggve, takes advantage of the Nordic registries and aims to give updated and valid estimates of cancer risk in IBD.

The study is part of the Tryggve use case programme. In order to learn more about it, we asked the I-SCAN researcher **Lise M. Helsingen** from the University of Oslo to tell us what the study is about.

A study that aims to reduce unnecessary clinical measures

I-SCAN stands for Inflammatory bowel disease - Scandinavian CANcer in IBD study. I-SCAN is a study exploring the risk of cancer in patients with inflammatory bowel disease, IBD, which is a chronic disease that often starts at a young age. IBD causes inflammatory changes in the gastrointestinal tract, and the most common symptoms are abdominal pain and bloody diarrhea. The disease can be very debilitating for the individual. Due to chronic inflammation of the gut, there has been a general belief that these patients have an increased risk of cancer, especially cancer of the large bowel.

Due to the perceived high cancer risk, these patients are often recommended regular colonoscopy surveillance, i.e.

Future

Challenges for Tryggve2 in 2020

- Consolidate the achievements of D1, D2/3 into a coherent (vision for a) Nordic infrastructure platform

- Produce concrete outputs, proof of concepts, demonstrators, services and use case implementations

- Disseminate the results actively

Challenges for Tryggve2 in 2020

Tryggve — 25.10.2019 — **neic**

D-2-3 Development roadmap for the secure systems within Tryggve, 2019

Abdulrahman Azab, Ali Syed, Antti Pursula, Henric Zazzi, Niclas Jareborg, Jonas Hagberg, and Juha Tornroos

Executive Summary

This document presents an overview of the current state and plans for the **architecture of the future sensitive data infrastructure** resulting from the Tryggve project, including secure data storage and cross-border processing components. More discussion is included on the treatment on cross-border processing of sensitive data, as this is the more experimental part of the project. The development of the sensitive data infrastructure is done in several deliverable teams of the project: Sensitive data archiving (D1), Production quality secure services (D2), Interoperability solutions (D3), Use cases (D4) and AAI (D5).

Tryggve operates a use case programme that engages with research teams to implement secure solutions for their research case. These use cases are driving the technology development in Tryggve by setting requirements and timelines for the development. The requirements from use cases include federated, pooled and on-demand processing. The proposed architecture for the Nordic sensitive data platform aims to be suitable for all these cross-border sensitive data use cases. In addition, the suggested platform is not limited to current use cases, nor even to Nordic projects, as all the development is done with the aim of generic solutions suitable for large-scale studies.

Tryggve

Tryggve2 D1 Product Planning

File   Edit   View   Insert   Format   Tools   Add-ons   Help   Last edit was made 5 days ago by blank dots

1. Tryggve D1 milestones and what we need in order to achieve them e.g. specific tryggve node requirements
   - **M9** 2020-04 Demonstrate stand-alone Sensitive Data Archive capability
     - Metadata minimal
     - Elixir AAI token integration for submission and data access
     - Data Access control
   - **M10** 2020-06 Technical capability to mirror metadata from CEGA to a local instance
     - Mechanisms to request, transfer and store metadata from CEGA to a local instance for datasets stored at the local instance
   - **M6** 2020-10 Pilot services operational in Nordic Nodes
     - Figure out if data needs to be publicly accessible - depends on the agreements with EGA
2. Use cases that we need to plan for in development
   - NorTwinCan
     - Support for deploying the solution
   - THL Biobank - *pending a meeting*
   - ***Use cases should be aligned with the deliverables and the EGA production***
3. moving to EGA pre-production - end of February
   - Elixir AAI integration for ingestion and outgestion
     - Mapping of EGA to Elixir IDs - dependent on EBI - *pending email response*

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
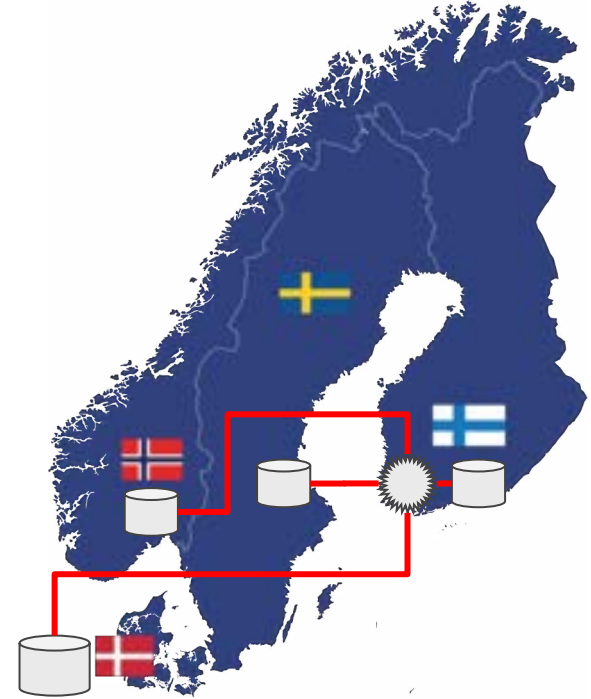
Beyond 2020...

Work in progress to prepare "Tryggve3" proposal to NeIC funding call

- Very tentative draft for work package structure
  - WP1: Federated Nordic Health/Biomedical Data Access (EGA technology uptake)
  - WP2: Sharing and managing geno/pheno datasets
  - WP3: Global/international standards in Nordics (to demonstrate cross-border interoperability)
  - WP4: Demonstrators/Use Cases (Higher level: 1M genomes, data for AI /ML, registry data...)
  - WP5: Impact goalsetting / Stakeholders / Sustainability
- Current partnership + ELIXIR Estonia
- Tentative budget: about half of Tryggve2
- Tentative duration 2-3 years

⊗neic
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": consolidation of Nordic sensitive data infrastructure (repositories and processing), prepare sustained operation **?**

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: Juha Törnroos (CSC)
juha.tornroos@csc.fi

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

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

Contact: tryggve@neic.no