Juni Palmgren
KOR Copenhagen
Nov 28, 2018

# Nordic Commons for Health Data

## Challenges and opportunities

NordForsk Programmes involving Nordic registers and biobanks

The Norwegian Presidency Project 2017: "Norden i omstilling"
Ministry of Health and Care Services, Norway

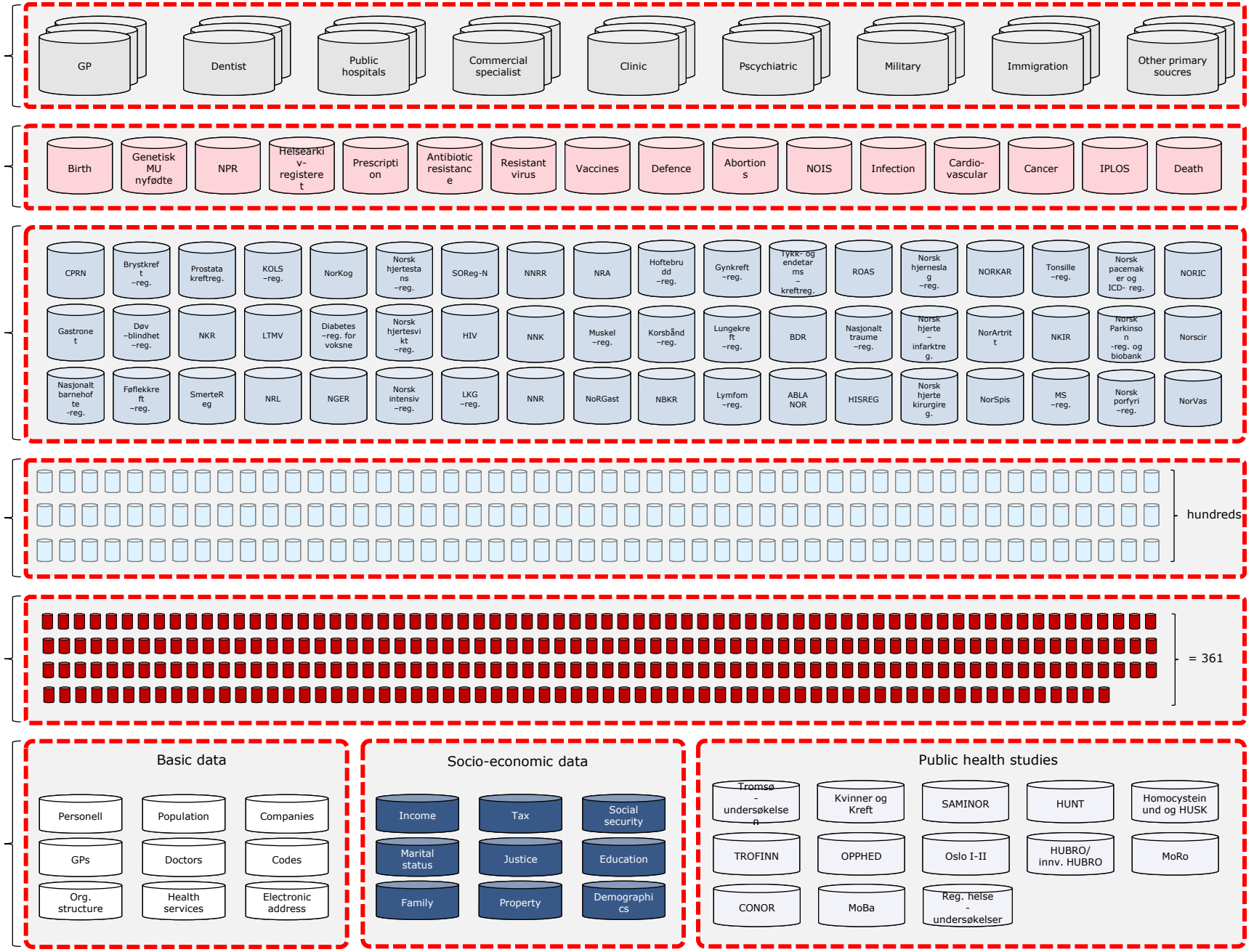
Nordic Council of Ministers EK-S →
NordForsk Programme for Health and Welfare

NordForsk

# Nordic registers and biobanks – A goldmine for resarch

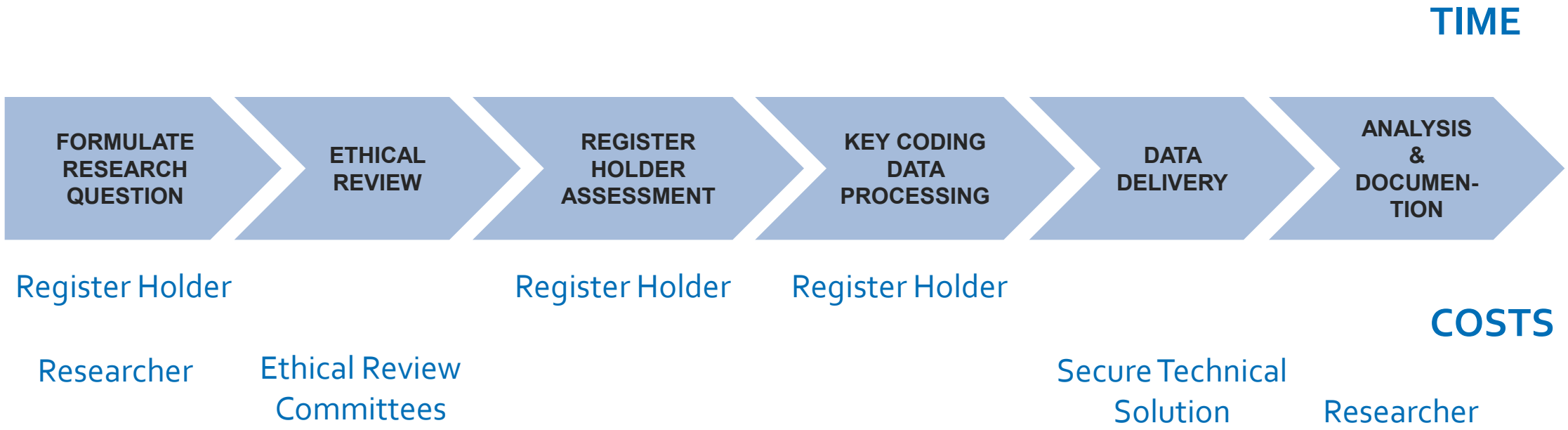
- The Nordic countries have a unique knowledge resource in its **longitudinal** disease and population registers
- The **Personal Identification Number** (PIN) allows linkage to biobanks and other databases
- Nordic data allows **comparisons** between countries or **joint analysis** at the Nordic level
- Population of 27 million - analysis at Nordic level will increase the possibility to find associations for **rare diagnoses and events**
- Could be used to find solutions to societal and public health challenges, evidence-based decisions, as well as to follow up political incentives

# Broad scope: Health data sources in Norway




From eHelsedir.

# The data retrieval process: key steps and actors



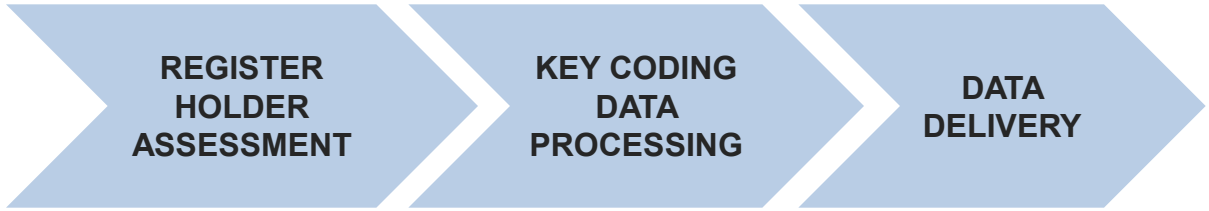

- **The process is similar in the Nordic countries**
- **The time and costs vary - increase with a Nordic study**

FORMULATE RESEARCH QUESTION → ETHICAL REVIEW

# Challenge: Lack of information

- What data exist? In which registers? Meaning?
- Can data in different registers be linked?
- How do I request data? Do I need any specific permissions?
- What legislation is applicable?

- **Limited knowledge of what permits are needed and what procedures are in place, even among researchers in the field**
- **Long and time-consuming dialogues with the register holders**
- **Need to apply for ethical permissions and data in each individual country**




# Challenge: Logistics

- Lack of coordinated processess between register holders both within countries and between Nordic countries
- Each national Statistics Agency inclined to lock-up their data at their own facility
- Need for secure technical environments for joint analysis

- **The researcher is dependent on a good dialogue between the register holders**
- **The researcher often needs to facilitate the process**
- **The register holders need to trust each other's technical platforms**

# The Nordic Political perspective

The Könberg Report

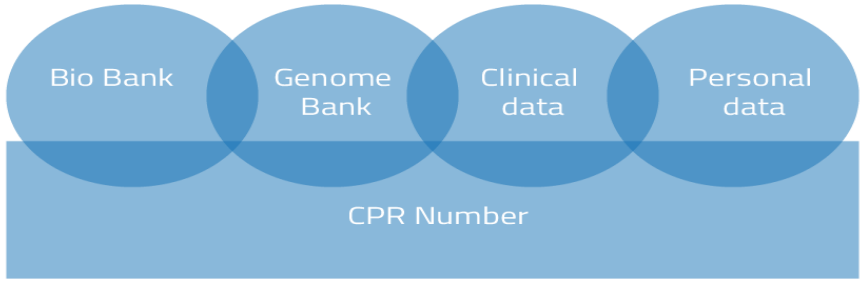
- Health cooperation is high on the Nordic political agenda
- Könberg Report - Chapter 4: Registers and Biobanks
- The Norwegian Presidency Project 2017: "Norden i omstilling" – focus on Nordic cooperation on health data and clinical trials
  - Ethical Review
  - Health Data
- National initiatives to promote health data utilisation

# National programs for Integrated Health Data

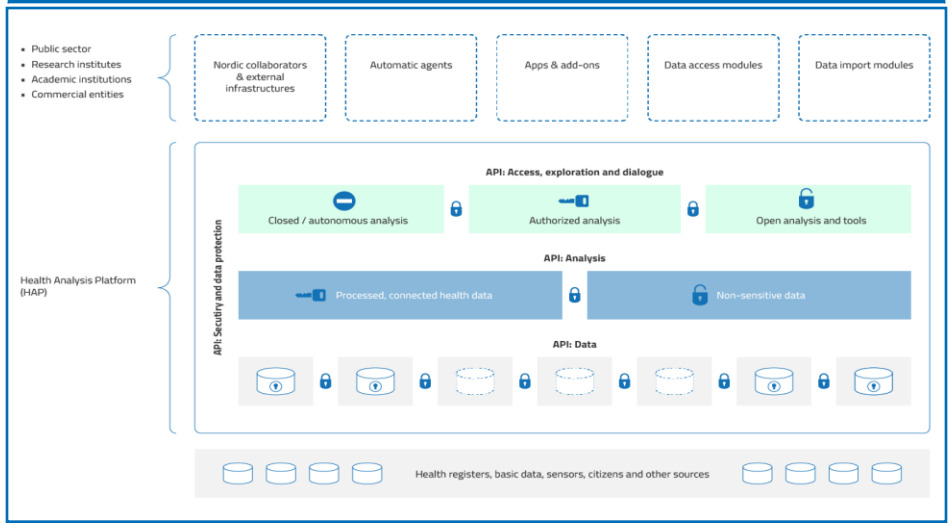
## National organizational, legal, financial and ethical perspective

## Focus on research, health care and industry

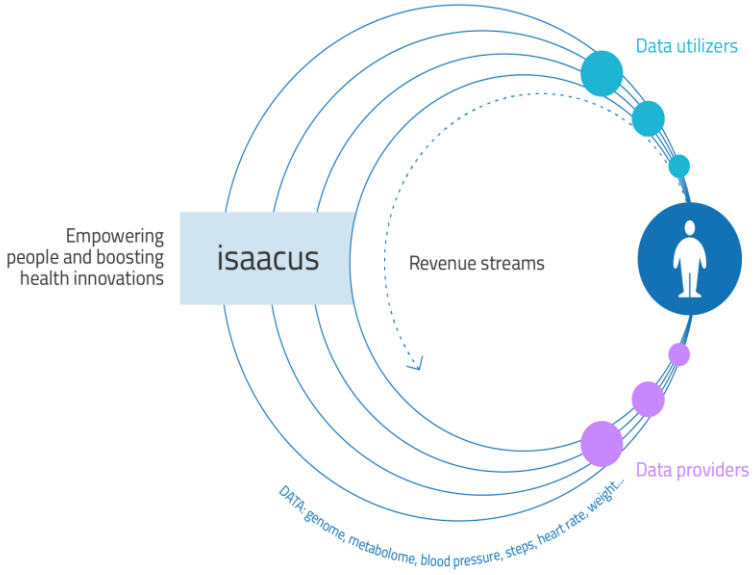
### Denmark - Unique position with Integrated Health Data




### Norway - The Norwegian Health Data Program is working on concepts for a national health analysis platform




### Finland - Isaacus programme




### Sweden – No specific national health data program

To date, there is no specific national health data program for Sweden. The landscape is rather fragmented.

Vetenskapsrådet has a **Register Infrastructure Programme** with a **RUT** data interface. Vinnova has a strategic innovation program **SweLife** and a recent initiative **Genomic Medicine Sweden**.

# Nordic Commons - vision

"A shared virtual space where scientists can work with the digital objects of **biomedical** research.

This is a system that will allow investigators to find, manage, share, use and reuse data, software, metadata and workflows".

# Components of a Commons eco-system

- **A computing environment,** such as the cloud and/or HPC (High Performance Computing) resources, which support access, utilization and storage of digital objects.

- **Data & metadata sets** that adhere to a set of Digital Object Compliance Principles which describe the properties of digital objects that enables them to be findable, accessible, interoperable and reusable (FAIR).

- **Software services and tools that enable;**
  - Scalable provisioning of compute resources.
  - Interoperability between digital objects within the Commons.
  - Indexing and thus discoverability of digital objects.
  - Sharing of digital objects between individuals or groups.
  - Access to and deployment of scientific analysis tools and pipeline workflows.
  - Connectivity with other repositories, registries and resources that support scholarly research.

# Towards a Nordic Commons for Health Data

- Nordic working groups on
  - **TECHNICAL SOLUTIONS:** Synchronizing national e-infrastructures for secure federated storage, sharing and analyses of sensitive personal data
  - **METADATA:** Focus on how to describe Nordic health data according to the FAIR* principles
  - **LEGAL FRAMEWORK:** Focus on legal questions related to technical solutions in 1.

**INTERIM REPORT TO BE SENT OUT FOR FACTUAL CHECKS JAN 2019**

**FINAL REPORT TO BE PRESENTED TO EK-S SPRING/MID 2019**

* Data being **F**indable-**A**ccessible-**I**nteroperable-**R**e-usable **(FAIR)**

# Technical solutions

## A Nordic secure orchestrator

*Working group*

- Peter Løngreen, Danish Technical University DTU, DK (Chair)
- Ali Syed, Danish Technical University DTU, DK
- Antti Pursula, Nordic e-Infrastructure Cooperation, FI
- Tommi Nyrönen, CSC, Elixir Finland, FI
- Hanne Cecilie Otterdal, Helsedataplattformen, NO
- Maria Francesca Lozzi, SIGMA2, NO
- Ann-Charlotte Sonnhammer, SNIC Uppsala University, SE
- Hanifeh Khayerri, Swedish Research Council, SE

NordForsk

## Current status

# National e-Infrastructures for Sensitive Personal Data

Denmark
### Computerome

**DeiC - National Life Science Supercomputer**: Computerome is the National dedicated e-infrastructure for health care and life sciences. It supports 1600 users locally and on European scene through its involvement in the ELIXIR and initiatives NeIC Tryggve. It provides a secure cloud service.

Finland
### CSC ePouta

**CSC ePouta** is a Finnish cloud computing environment delivered as Iaas (*Infrastructure as a Service*) designed for processing sensitive data. The ePouta cloud is being routinely used by several user groups, including national Center of Excellence for Tumor Genetics and Finnish Institute for Molecular Medicine.

Norway
### TSD

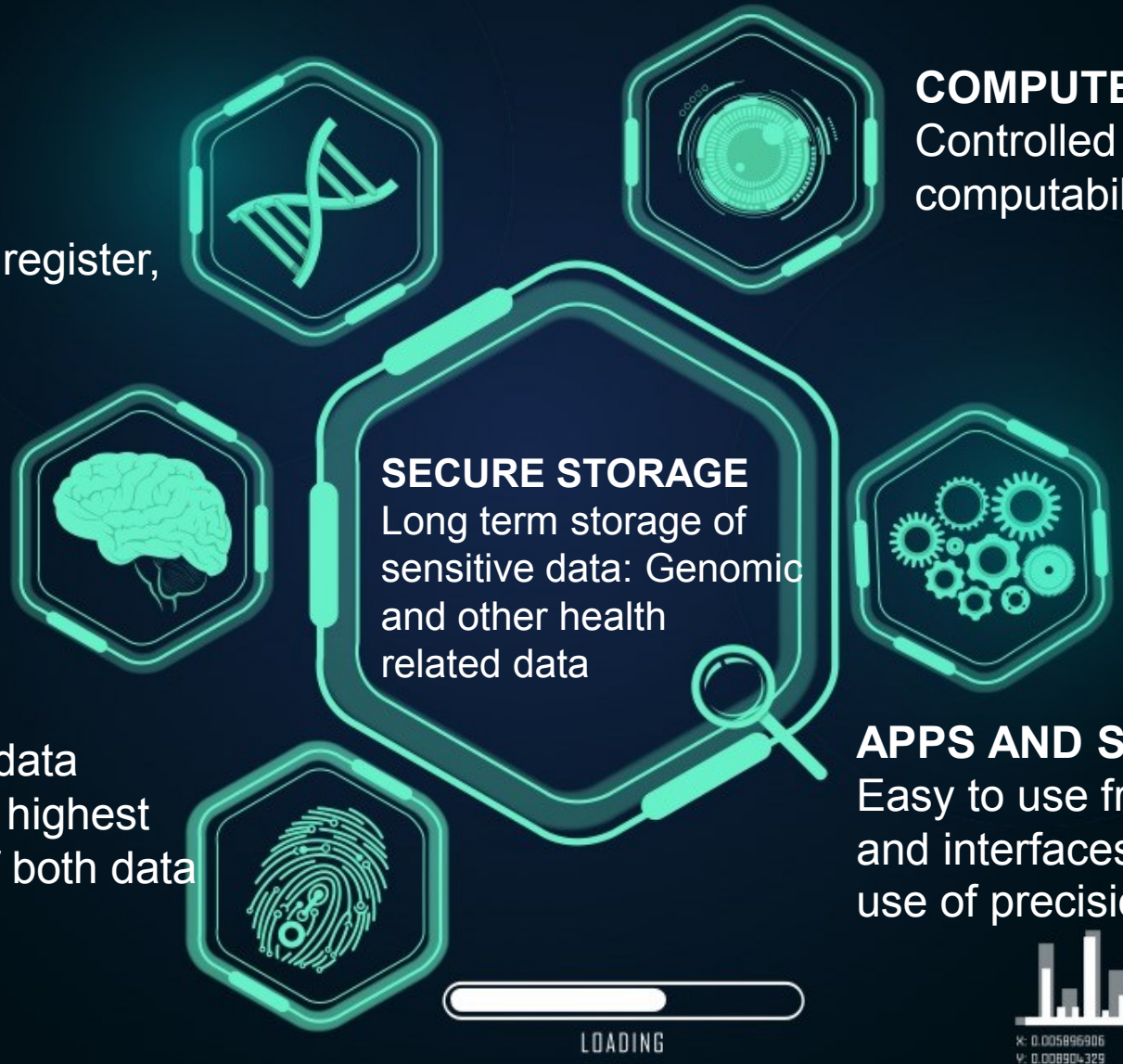
**The project Services for Sensitive Data (TSD)**, initiated by USIT (The University Centre of Information Technology) at The University of Oslo, is a national service to researchers in Norway and abroad for storing and processing sensitive data, including health data. TSD provides a secure cloud service in production environment.

Sweden
### Bianca, Mosler, RUT and MONA

Currently no unified national cloud solution for health and welfare, but several actors are involved offering their own local solutions to health and welfare data producers and users. However, the e-infrastructures for sensitive *research data* are in the forefront and are best qualified to be considered national cloud solutions. These would be the:
**Bianca system on the Swedish National Infrastructure for Computing**
**Swedish ELIXIR system Mosler**
**Swedish Registry Utilizer Tool being built (RUT)**
**Statistics Sweden's Microdata Online system (MONA)**

Collaboration through the Tryggve/Tryggve2 (2014-2020) projects for sensitive data hosted by the Nordic eInfrastructure Collaboration NeIC

# NORDIC TECHNICAL SOLUTION
## BUILDS ON EXISTING COMPONENTS




**INTEGRATION**
patient, clinical, register, research data

**COMPUTE POWER**
Controlled access and computability of data

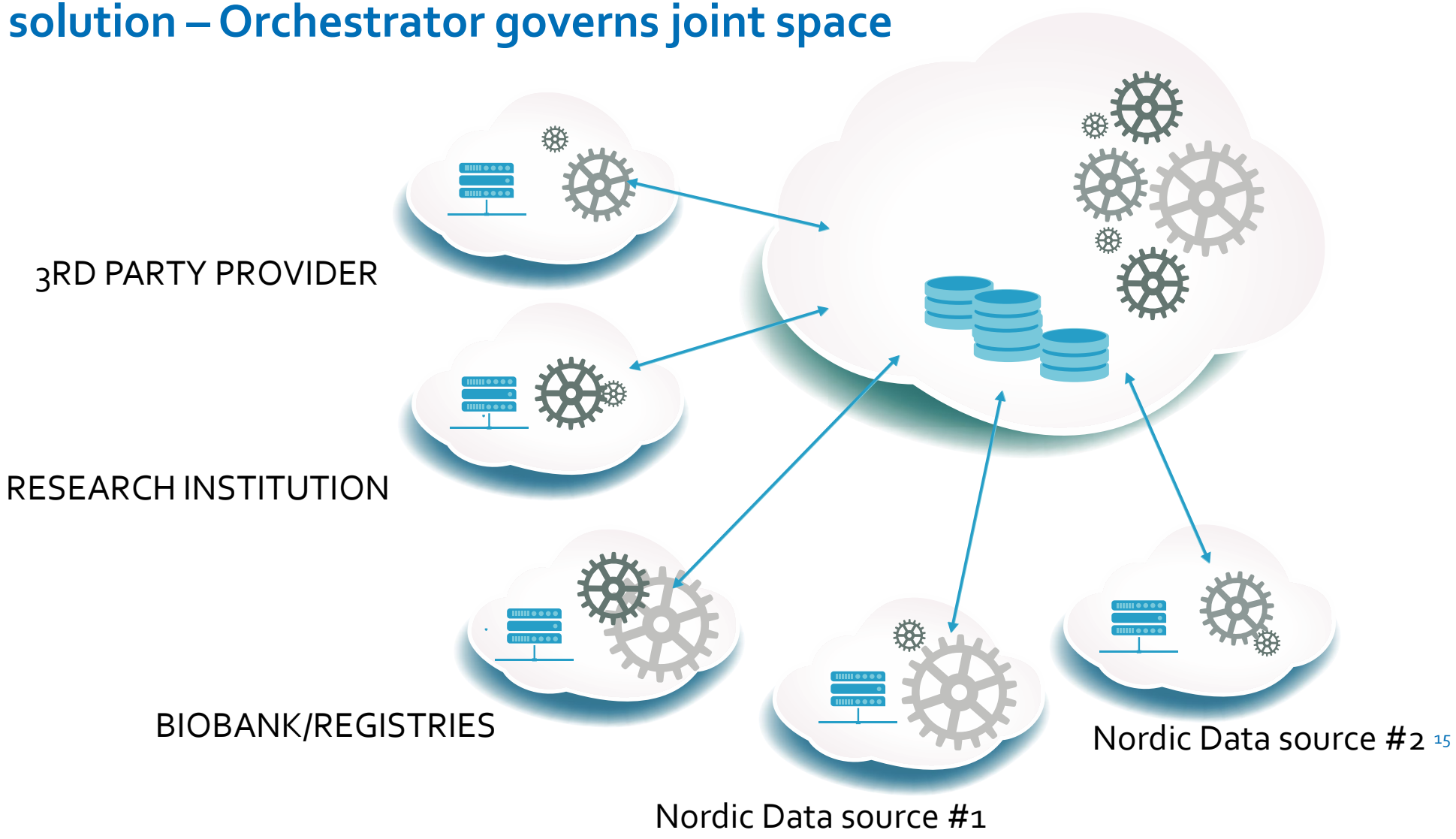
**SECURE STORAGE**
Long term storage of sensitive data: Genomic and other health related data

**SECURE ACCES**
Prevent abuse of data by introducing the highest level of security of both data and connections

**APPS AND SERVICES**
Easy to use front-end apps and interfaces for clinical use of precision medicine

# Example flow

## Federated solution – Orchestrator governs joint space



# Automated log – a Nordic Log Store

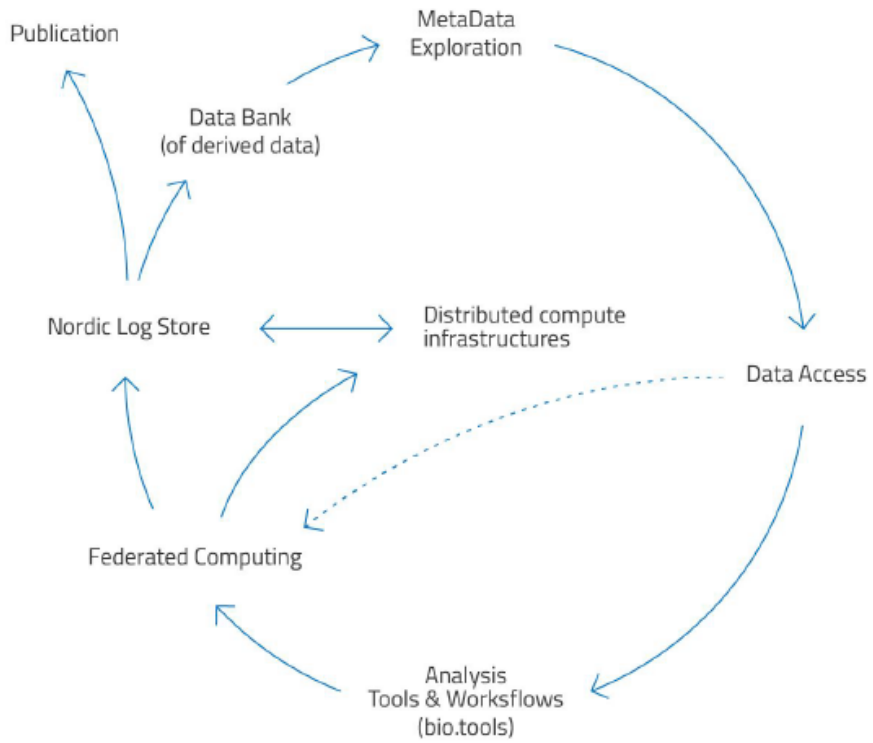



Fig: Life cycle of data in a project running within the Nordic Secure Cloud

- The Orchestrator distributes the tasks to the available data centers.
- All operations are logged in a Nordic Log Store.
- The derived data are assigned an identifier (e.g. DOI).
- Metadata of the analysis process feeds into the original metadata repositories.
- The loop is closed!

# Nordic health metadata

*Working group*

- Magnus Eriksson, Swedish Research Council, SE (Chair)
- Jeppe Klok Due, Det koordinerande organ för registerforskning, KOR, DK
- Arto Vuori, National Institute for Health and Welfare, FI
- Truls Korsgaard, Directorate for e-Health, NO

NordForsk

# Describe data

## Findable

Good descriptions of the data we want to find, relevant attributes on appropriate levels

- "Rich metadata"

Accessible from a solution providing search functionality

- "Indexed in a Searchable resource"

Be able to handle same names on datasets, variables, researchers, publications…etc without mixing them up.

- "Persistent identifier"

## Accessible

So we can evaluate and find it again, reuse it in different contexts when appropriate.

- "Metadata are accessible, even when the data are no longer available."

Easy to access the metadata using software.

- "(meta)data are retrievable by their identifier using a standardized communications protocol."

For those who have permissions

- "the protocol allows for an authentication and authorization"

Without needing to use vendor specific software or solutions in order to be able to access

- "the protocol is open, free, and universally implementable."

Based on FORCE11 - https://www.force11.org/group/fairgroup/fairprinciples

# Describe data

## Interoperable

The meaning of the data are decribed in a way that provides context and make it understandable not only by people but also computers.

- ”(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation”

The semantics (concepts and concept systems) are described with references to terminologies, ontologies etc.

- ”(meta)data use vocabularies that follow FAIR principles.”

## Reusable

Detailed descriptions of the data content we want to reuse by relevant attributes

- ”meta(data) have a plurality of accurate and relevant attributes.”

Making sure we know in what way we are allowed to use it.

- ”(meta)data are released with a clear and accessible data usage license.”

And with detailed descriptions of how it has been produced, from which sources, by whom using what resources.

- ”(meta)data are associated with their provenance.”

In a way that it can be easily used with tools and in combination with other data from the domain.

- ”(meta)data meet domain-relevant community standards.”

Based on FORCE11 - https://www.force11.org/group/fairgroup/fairprinciples

# Levels of detail: Metadata & Semantics

| Descriptions of | Content | Examples | Ex. Standards |
|---|---|---|---|
| Framework standards | How to describe data and concepts used for descriptions | Concept, ConceptSystem, Variable, Population, Dataset… | ISO11179, GSIM |
| Dataset level standards | Attributes to describe the dataset. | Creator, Title, Publisher, Publication year, ResourceType, Funding… (DataCite) | DataCite<br>DDI<br>DCAT-AP<br>… |
| Domain specific standards | What should be described and details on how. | Patient (resource, domain, unittype…)<br>• Birth time (attribute, variable…)<br>• Nationality<br>Organisation<br>• Alias<br>• Period…<br>Medication (HL7 FHIR) | HL7 FHIR<br>HL7 V3<br>DDI<br>OMOP |
| **Related standards** | | | |
| Semantics | Concepts and terms to define meaning and context for humans and computers. | Läkemedel<br>• "SCTID: 410942007, Läkemedel" (SnomedCT)<br>• "Läkemedel för humant eller veterinärt bruk, i sin bruksfärdiga form. Hit räknas också de ämnen som används i framställningen av den färdiga preparatformen." (Mesh) | SnomedCT<br>Mesh<br>Loinc<br>Nationellt fackspråk |
| Persistent Identifiers | Unique keys for metadata and data resources. | Persistent Identifiers for researchers, Data Sets… | DOI<br>ORCHID |

# Status in the Nordics – rough estimate (Fall-2018)

| Domain | Country | Findable | Accesible | Interoperable | Reusable |
|---|---|---|---|---|---|
| Health Registers | | | | | |
| Registries of Clinical Quality | | | | | |
| Biobanks | | | | | |
| OMICS | | | | | |
| Laboratory data | | | | | |
| Health Surveys/Cohort studies | | | | | |
| Socioeconomic registries | | | | | |

# Legal Framework

*Working group:*

- Marjut Salokannel, FI
- Victoria Söderqvist, SE
- Manolis Nymark, SE
- Ragnhild Angell Holst, NO
- Lars Emde Poulsen, DK

NordForsk

# 3. Legal Framework

- Ensure accreditation of compute facilities
- Ensure Nordic alignment of national safeguards in the wake of GDPR
- Anchor the Nordic solution with national data protection authorities
- Ensure certification and set up a code of conduct for the Nordic solution

# Thank you for your attention!

**Juni Palmgren**
Karolinska Institutet, Stockholm
Chair of NordForsk Expert group on Health Data Infrastructure

Coordinator
**Maria Nilsson,** Special Adviser
Leader of the NordForsk Programme on Health and Welfare
**maria.nilsson@nordforsk.org**