HEALTH
DATA HUB

Health data access and data governance structures

International benchmark 2020

Presentation outline

1. Updates on the European context of health data governance
2. International benchmarking method and overview of analyzed structures
3. Results by category of analysis for the structures directly comparable to the Health Data Hub
4. Best practices from other actors within the international health data ecosystem

1. Updates on the European context of health data governance

Insufficient data sharing and data re-use for research and innovation in the European Union

Limited data access and availability for secondary use

- Lack of clarity on the re-use of public and private sector data for public interest research
- Lack of visibility and difficulties to access data
- Scarcity of ‘voluntary’ mechanisms to provide and exchange data

Few European solutions for data storage and data processing

- Questions around data volume, security, data analytics tools, data protection…
- Discussions around GAIA-X and the European Health Data Space
- Roadmap for a European Health Data Space

Fragmentation of the European single market

The absence of a global approach to data governance

- Legal obstacles : Heterogeneous implementation of the GDPR
- Technical obstacles between sectors (ex: standardisation and interoperability) : the majority of databases is still not interoperable

Lack of skills in health data management and a culture of health data sharing

Lack of citizens’ *empowerment* mechanisms around health data

In response to these challenges, the European Commission has launched a large scale initiative that aims at creating the future European Health Data Space through various legislative instruments and funding programmes

Source: : From an official presentation of the European Commission in May 2020

In response, the Commission has started to develop instruments for the future European Health Data Space

European data strategy	Communication proposed by the European Commission in February 2020
Data Governance Act proposition	First version proposed by the European Commission end of 2020, Member States and EU-level discussions in 2021 and potential adoption early 2022
Legislative proposal for the European Health Data Space	Will be proposed by the European Commission in the first half of 2022 and will be based on public consultations conducted throughout 2021
TEHDaS Joint Action	25 Member States work towards options for governance, infrastructure, data quality, 'data altruism' and citizens' involvement in the European Health Data Space
Other investments to support the European Health Data Space	As part of the 2021-2027 EU4Health funding programme and the common data spaces, as well as Horizon and Digital Europe programmes on digital health and secondary use of health data
Codes of conduct by sector	Codes of conduct to support the secondary use of health data developed jointly by relevant actors
European Health Data Space Pilot Project	Proof of concept to assess the feasibility of the implementation of a small-scale European Health Data Space through collaboration of different nodes, including the HDH, Findata, the Danish, Norwegian Health Data Authorities, EMA, ECDC, Elixir, BBMRI… and the European Commission, to be launched in the second quarter of 2022
Other European financing opportunities	eHealth assessments, studies and call for projects for Member States to facilitate the economic recovery and resilience (Recovery and Resilience Facility), the European Regional Development Fund, the European Social Fund, InvestEU…

EHDS: What are the problems ?

Use of health data for healthcare (primary)

Sharing of health data for healthcare

- **Limited control to and access of patients to their data**
- **Unnecessary health care** (overutilization, overuse, overtreatment) leading to high costs
- **Insufficient exchange** of health data
- **Limited (cross-border) interoperability** between healthcare providers

Single market for digital health products and services

- Non-uniform national **legislative frameworks**
- Uneven **quality / interoperability** frameworks
- Uneven **procedures** for prescriptions, reimbursement, liability

Re-use of health data (secondary)

Access to health data for research, innovation, and public health policy making

- **Low re-use** of health data
- **Complex cross-border access** to health data
- Fragmented digital **infrastructures**

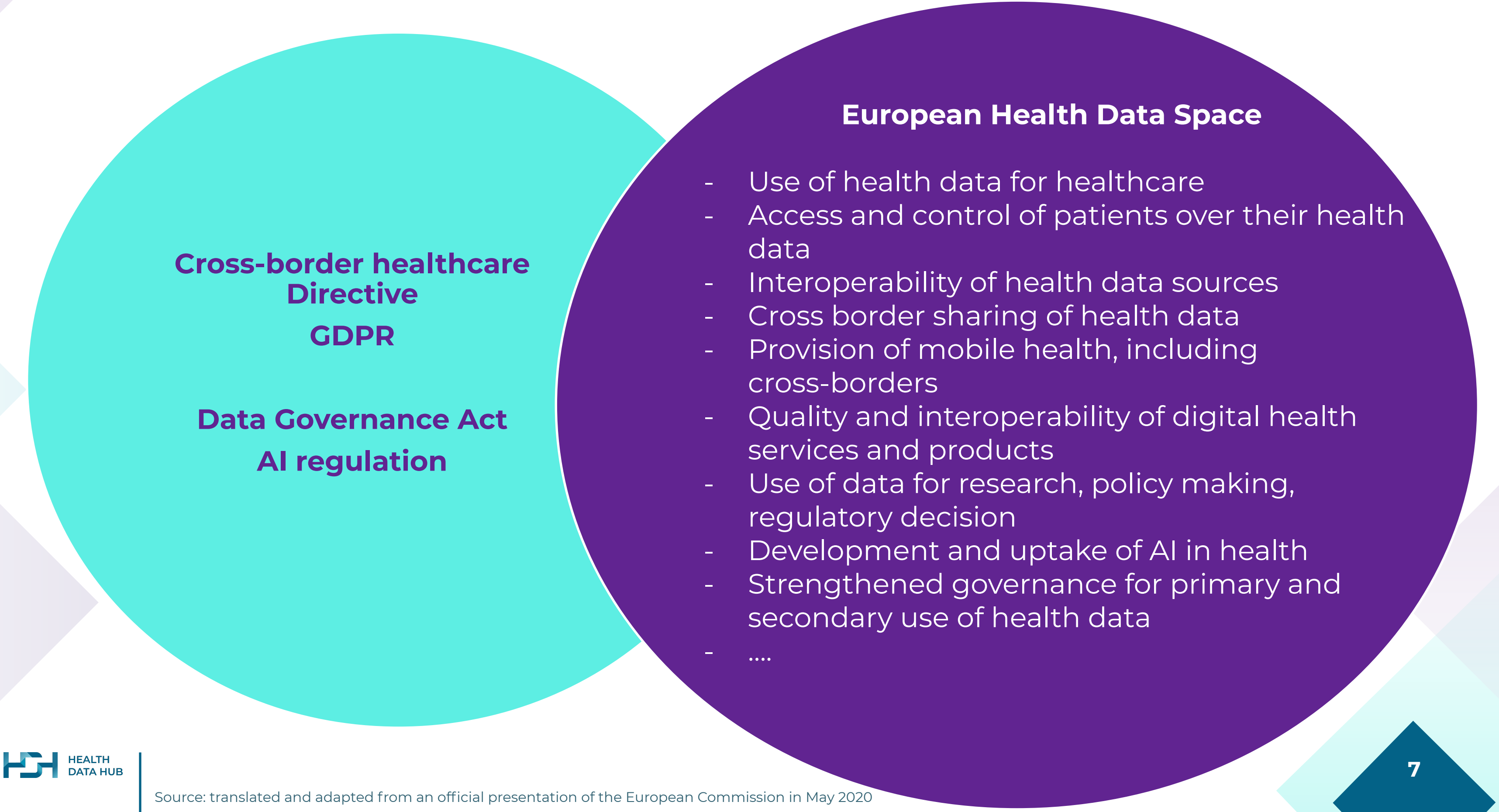
Artificial Intelligence

- Limited **provision** of data for training of AI
- Difficulties for regulators to evaluate
- **AI algorithms**
- Uncertainty on AI **liability** in health

HEALTH DATA HUB

Source: : From an official presentation of the European Commission in May 2020

EHDS: Articulation with EU regulatory framework



Cross-border healthcare Directive

GDPR

Data Governance Act

AI regulation

European Health Data Space

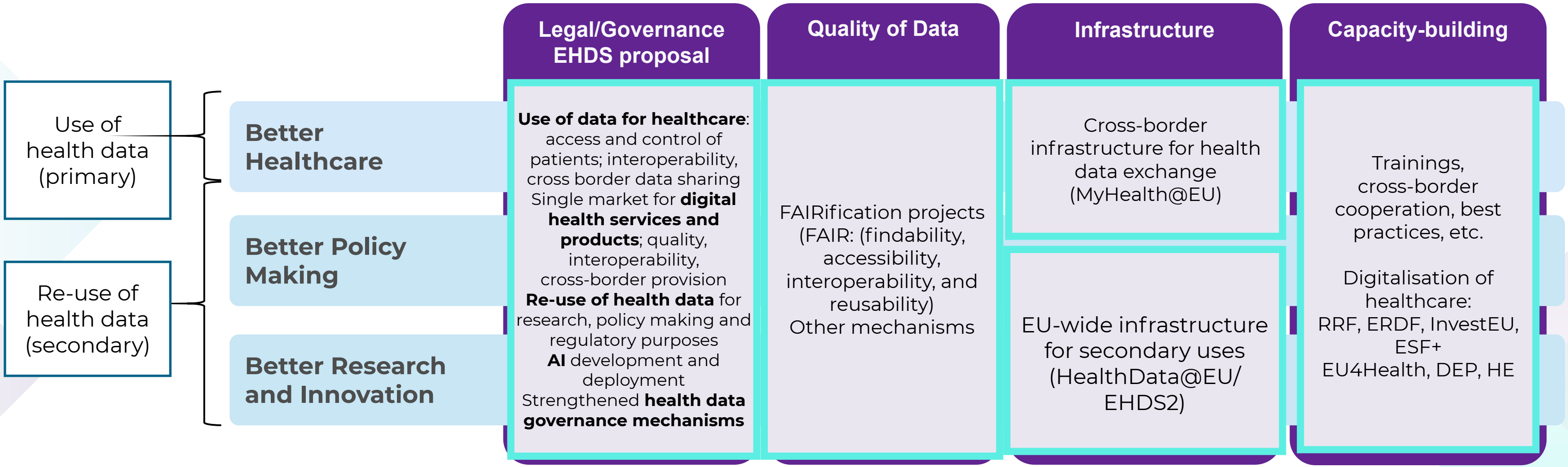
- Use of health data for healthcare
- Access and control of patients over their health data
- Interoperability of health data sources
- Cross border sharing of health data
- Provision of mobile health, including cross-borders
- Quality and interoperability of digital health services and products
- Use of data for research, policy making, regulatory decision
- Development and uptake of AI in health
- Strengthened governance for primary and secondary use of health data
- ....

Source: translated and adapted from an official presentation of the European Commission in May 2020

Primary and secondary uses of health data

Timely and simplified *exchange of* and *access to* health data for different use cases:

- Healthcare provision, access and control of patient over their data, (cross-border) exchange of health data
- Provision of digital health services (including telehealth and m-health)
- Research (e.g. on cancer, rare diseases, COVID-19, etc.), pharmacovigilance, public health, policy making



Use of health data (primary)

Re-use of health data (secondary)

Better Healthcare

Better Policy Making

Better Research and Innovation

Legal/Governance EHDS proposal

Use of data for healthcare: access and control of patients; interoperability, cross border data sharing
Single market for **digital health services and products**; quality, interoperability, cross-border provision
Re-use of health data for research, policy making and regulatory purposes
AI development and deployment
Strengthened **health data governance mechanisms**

Quality of Data

FAIRification projects (FAIR: (findability, accessibility, interoperability, and reusability)
Other mechanisms

Infrastructure

Cross-border infrastructure for health data exchange (MyHealth@EU)

EU-wide infrastructure for secondary uses (HealthData@EU/ EHDS2)

Capacity-building

Trainings, cross-border cooperation, best practices, etc.

Digitalisation of healthcare:
RRF, ERDF, InvestEU, ESF+
EU4Health, DEP, HE

Focus on the joint action TEHDaS

Launched in February 2021, the **Joint Action Towards a European Health Data Space (TEHDaS)**, with input from 25 EU Members States, seeks to **develop and promote legal, technical and infrastructure concepts for the sharing of health data for secondary use**, with the goal to improve health research and innovation in Europe and public health

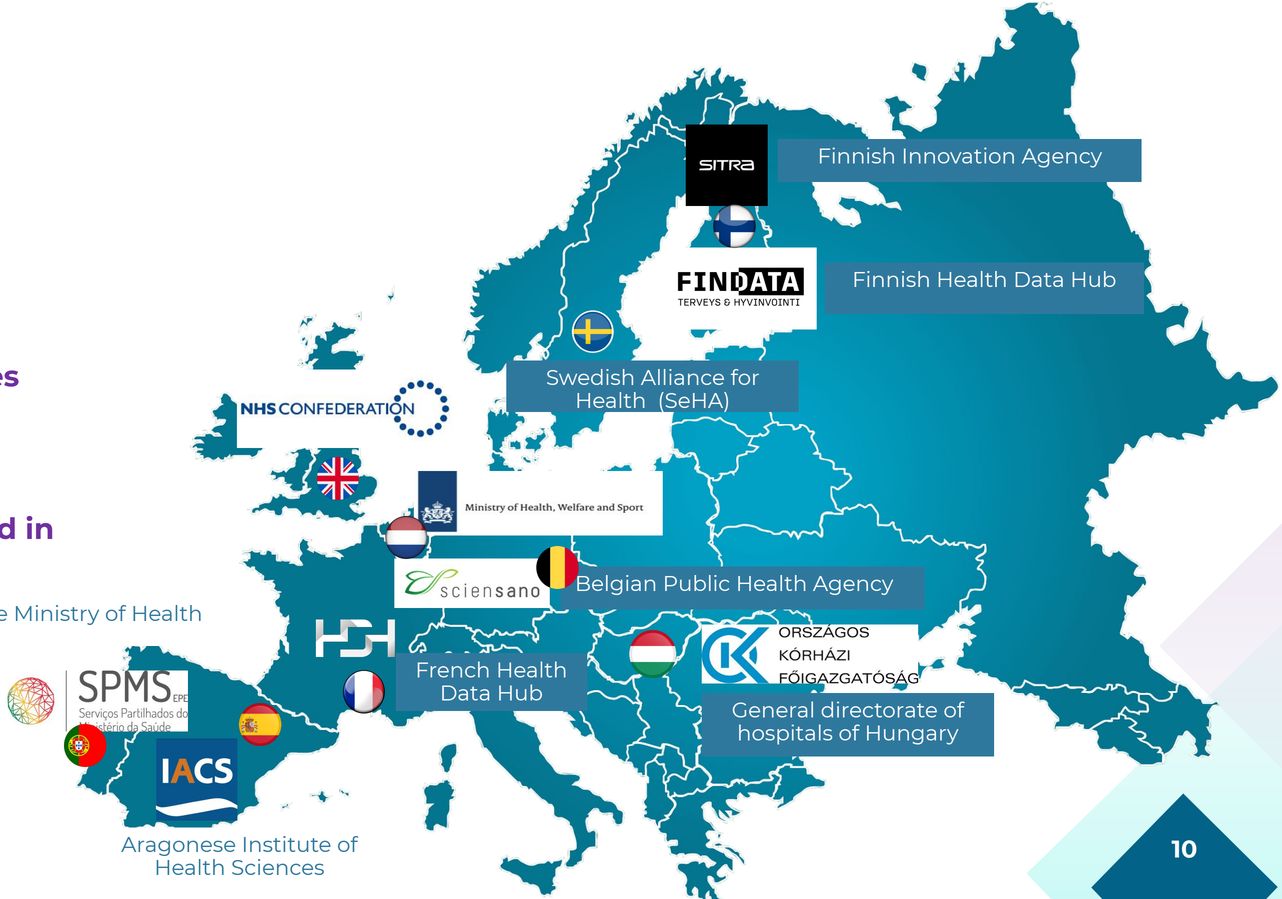
Vision of the European Health Data Space:
Citizens, communities and companies should benefit from **protected and secure access to interoperable health data all around Europe**

From mid-2021 onwards, **as part of TEHDaS work packages, recommendations will be issued** to inform the European Commission on the creation of the European Health Data Space, including the proposition of **a European law for the European Health Data Space (expected early 2022)**

Key stakeholders of TEHDaS

4 Work Packages

12 french stakeholders mobilised in TEHDaS



Finnish Innovation Agency

Finnish Health Data Hub

Swedish Alliance for Health (SeHA)

Belgian Public Health Agency

Portuguese Ministry of Health

French Health Data Hub

General directorate of hospitals of Hungary

Aragonese Institute of Health Sciences

The HDH participates actively and coordinates the contributions of other French stakeholders

The HDH, appointed as the competent authority for France and collaborates with its affiliated entities

AGENCE
DU NUMÉRIQUE
EN SANTÉ

Hospices Civils de Lyon

Inserm

Aix*Marseille
université
Initiative d'excellence

UNIVERSITÉ
TOULOUSE III
PAUL SABATIER
Université
de Toulouse

An excellent opportunity to promote the work at European scale

France will take over the Presidency of the EU Council in the first semester of 2022, from 1 January to 30 June.

- A presentation of first results will be possible during several key events, such as the high-level conference on "Citizenship, Ethics and Health Data", planned for February 2022 by the French, to give visibility to the numerous initiatives on health data governance.

4 Work Packages focused on coordination and communication

Coordination and implementation of the project with a dedicated project management office, distribution of funding

WP1 Coordination

WP3 Evaluation

Project monitoring and evaluation. Define and track progress indicators and ensure that the project is reaching its objectives.

Communicating results and activities of all WPs. Sending a newsletter, create and update TEHDaS website, social media....

WP2 Communication

WP4 Outreach, engagement and sustainability

Onboarding **stakeholders**, reflecting on how to make the project sustainable in the long run

The Health Data Hub involved in particular in Work Packages 5 on health data governance and WP8 on citizen engagement within TEHDaS

WP5

Governance

Propose recommendations for the **EHDS legislation and health data governance.** Set up use-cases: verify for example, how to access health data on different health data platforms in the EU

WP8

Data altruism

Identify common and best practices as well as case studies of RGPD compliant **data altruism**

HDH co-lead

Secondary use of health data

WP6

Data quality

Reflect on existing **health data standards** to determine how health information is organised at the national and European level. Derive recommendations at the European level.

WP8

Citizen engagement

Understand and consider **citizens' perceptions of health data and data sharing**

WP7

Infrastructure

Define the **technical solutions and services for the cross-border sharing and use of health data,** in a secure framework and respecting the rights of individuals

2. International benchmarking method and overview of analyzed structures

Results of the EU health support study report (2020)

Access mechanisms for the secondary use of health data	Number of Member States	Member States
Access is granted after authorisation by research ethics committee (REC) or data protection agency (DPA)	22	Belgium, Czech Republic, Denmark, Germany, Estonia, Ireland, Greece, Spain, France, Croatia, Italy, Cyprus, Lithuania, Latvia, Luxembourg, Hungary, Austria, Poland, Portugal, Romania, Finland, Sweden, UK
Data controller provides direct access without consultation of an ethics committee or a DPA	7	Denmark, Croatia, Italy, the Netherlands, Austria, Slovenia, Finland, UK
Some form of centralised governance body exists	13	Belgium, Denmark, Germany, Ireland, Greece, France, Cyprus, Malta, the Netherlands, Portugal, Slovenia, Slovakia, Finland, UK

Source: Assessment of the EU Member States' rules on health data in the light of GDPR, page 98
Link: https://ec.europa.eu/health/sites/health/files/ehealth/docs/ms_rules_health-data_en.pdf

International benchmarking criteria

Update of the analysis grid and interview structure

1 **Step 1:** Update of the categories used for the 2019 benchmark and inclusion of new analysis criteria, including ***Citizens*** and ***COVID-19 initiatives.***

2019

- Characteristics of the platform and services proposed
- Type of data
- Data access procedure
  - Pre-requisites
  - Actors that can access data
  - Value sharing (exclusivity period)
  - Academic competitiveness
- Approach and instructions of access files
- Data access governance (data sharing, transparency, ethics)
- Validation committee
- Formalization of application and review of requests
- Economic model
- Metadata

\+

2020 (new criteria)

- Strategic roadmap
- Scientific valorization
- Informations for citizens, their rights and implication
- Communication and external collaborations
- COVID-19 data centralization initiatives

2 **Step 2:** Identify and organize contact with international health data access and governance structures

3 **Step 3:** Consolidating the content of the interviews into reports; from which we deduce detailed data sheets by structure and a comparative analysis by theme

International benchmarking objectives

Objectives of the International Benchmark

An international Benchmark will allow the Health Data Hub and its partners to get an overview of the good international **practices for health data access and health data storage, the relations with data controllers and citizens,** the **economic models,** and the **scientific valorization rules**. Besides, the interviews are a great opportunity to create and develop an international network, source of potential synergies and prospective partnerships.

11 countries represented
17 structures benchmarked
14 interviews

An international network gathering > 1500 contacts

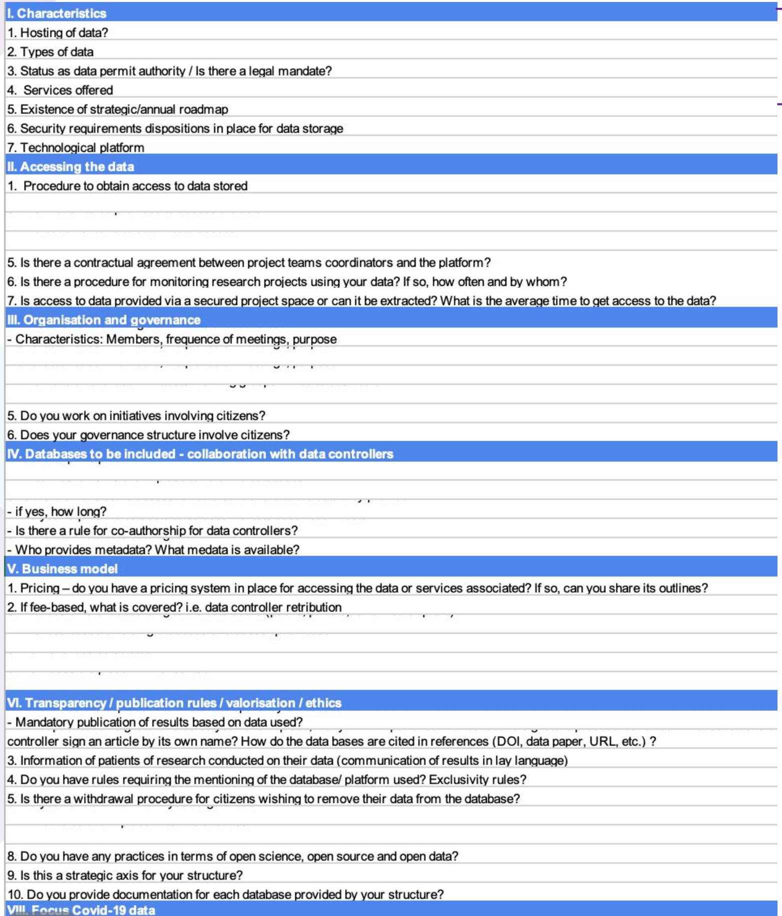
A publication in the making

Continuous update

11 countries & 17 international structures benchmarked

HEALTH DATA HUB

Overview of interview outline



I. Characteristics
1. Hosting of data?
2. Types of data
3. Status as data permit authority / Is there a legal mandate?
4. Services offered
5. Existence of strategic/annual roadmap
6. Security requirements dispositions in place for data storage
7. Technological platform

II. Accessing the data
1. Procedure to obtain access to data stored
5. Is there a contractual agreement between project teams coordinators and the platform?
6. Is there a procedure for monitoring research projects using your data? If so, how often and by whom?
7. Is access to data provided via a secured project space or can it be extracted? What is the average time to get access to the data?

III. Organisation and governance
- Characteristics: Members, frequence of meetings, purpose
5. Do you work on initiatives involving citizens?
6. Does your governance structure involve citizens?

IV. Databases to be included - collaboration with data controllers
- if yes, how long?
- Is there a rule for co-authorship for data controllers?
- Who provides metadata? What medata is available?

V. Business model
1. Pricing – do you have a pricing system in place for accessing the data or services associated? If so, can you share its outlines?
2. If fee-based, what is covered? i.e. data controller retribution

VI. Transparency / publication rules / valorisation / ethics
- Mandatory publication of results based on data used?

controller sign an article by its own name? How do the data bases are cited in references (DOI, data paper, URL, etc.) ?
3. Information of patients of research conducted on their data (communication of results in lay language)
4. Do you have rules requiring the mentioning of the database/ platform used? Exclusivity rules?
5. Is there a withdrawal procedure for citizens wishing to remove their data from the database?
8. Do you have any practices in terms of open science, open source and open data?
9. Is this a strategic axis for your structure?
10. Do you provide documentation for each database provided by your structure?

VIII. Focus Covid-19 data

Category / Topic addressed

Specific questions

International benchmarking method

Analysis and organization of results into categories

Interview reports were analyzed and two categories identified. **Category 1 structures hosting data** and **Category 2 structures not hosting data but impacting governance and quality of health data at the national and international level**. The structures were then analyzed according to the specific study topics.

Study topics

- **Platforms**
- **Scientific valorization**
- **Communication**
- **Citizens engagement and involvement**
- **COVID-19 data related initiatives**

1 **Observed practices and their frequency**

➔ **Common practice**

Is defined as a frequently observed practice common to more than 50% of platforms.

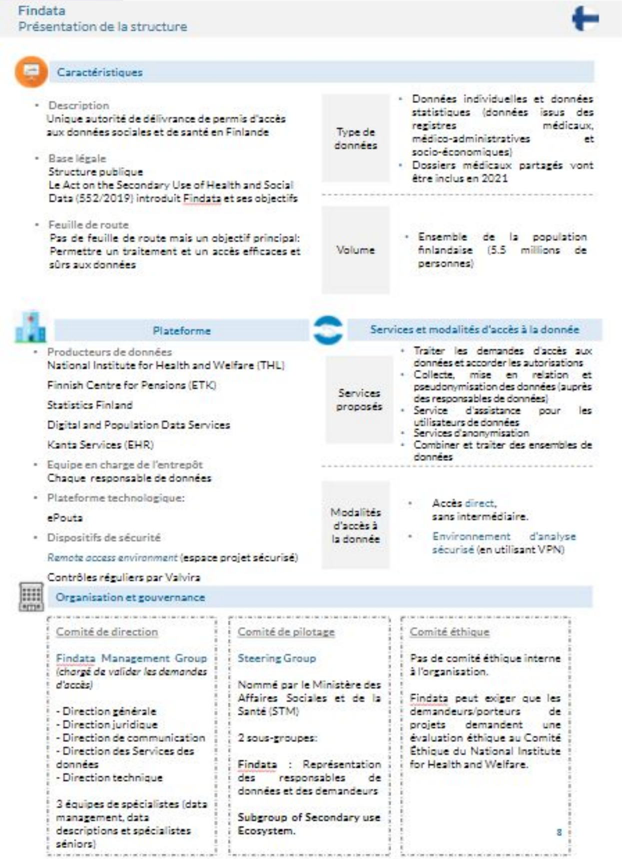
➔ **Ad hoc practice**

Defined as a practice observed less frequently (in <50% of observed structures).

2 **Highlight(s)**

An outstanding example or best practice in a given structure for each topic of study; highlighting whether **differing from general observations (unique practice)** or **with positive echo on international community**.

Example category 1 structure results presentation



Findata
Présentation de la structure

Caractéristiques

- Description
  Unique autorité de délivrance de permis d'accès aux données sociales et de santé en Finlande
- Base légale
  Structure publique
  Le Act on the Secondary Use of Health and Social Data (552/2019) introduit Findata et ses objectifs
- Feuille de route
  Pas de feuille de route mais un objectif principal: Permettre un traitement et un accès efficaces et sûrs aux données

Type de données	Données individuelles et données statistiques (données issus des registres médicaux, médico-administratives et socio-économiques) Dossiers médicaux partagés vont être inclus en 2021
Volume	Ensemble de la population finlandaise (5.5 millions de personnes)

Plateforme

- Producteurs de données
  National Institute for Health and Welfare (THL)
  Finnish Centre for Pensions (ETK)
  Statistics Finland
  Digital and Population Data Services
  Kanta Services (EHR)
- Equipe en charge de l'entrepôt
  Chaque responsable de données
- Plateforme technologique:
  ePouta
- Dispositifs de sécurité
  Remote access environment (espace projet sécurisé)
  Contrôles réguliers par Valvira

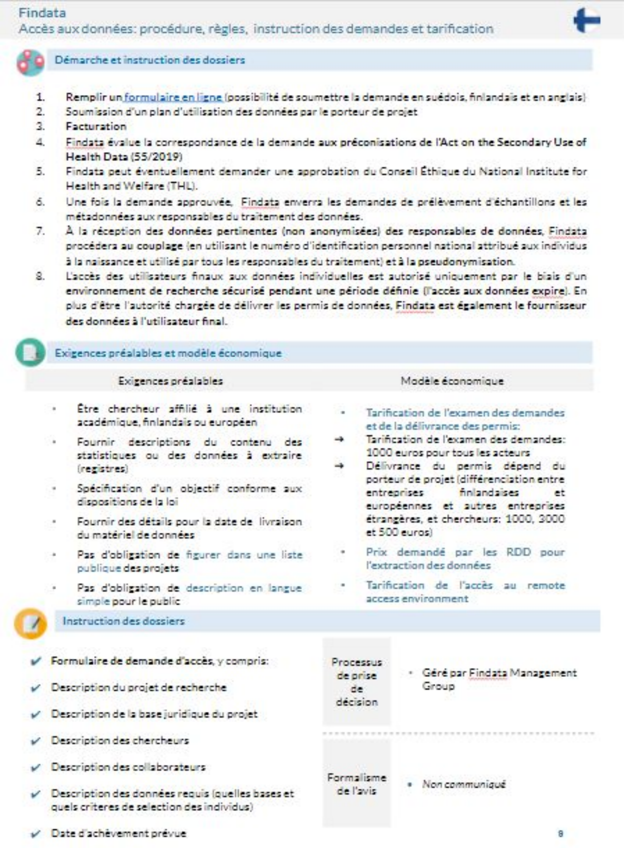
Services et modalités d'accès à la donnée

Services proposés	Traiter les demandes d'accès aux données et accorder les autorisations Collecte, mise en relation et pseudonymisation des données (auprès des responsables de données) Service d'assistance pour les utilisateurs de données Services d'anonymisation Combiner et traiter des ensembles de données
Modalités d'accès à la donnée	Accès direct, sans intermédiaire. Environnement d'analyse sécurisé (en utilisant VPN)

Organisation et gouvernance

Comité de direction	Comité de pilotage	Comité éthique
Findata Management Group *(chargé de valider les demandes d'accès)* - Direction générale - Direction juridique - Direction de communication - Direction des Services des données - Direction technique 3 équipes de spécialistes (data management, data descriptions et spécialistes séniors)	Steering Group Nommé par le Ministère des Affaires Sociales et de la Santé (STM) 2 sous-groupes: Findata : Représentation des responsables de données et des demandeurs Subgroup of Secondary use Ecosystem.	Pas de comité éthique interne à l'organisation. Findata peut exiger que les demandeurs/porteurs de projets demandent une évaluation éthique au Comité Éthique du National Institute for Health and Welfare.

8



Findata
Accès aux données: procédure, règles, instruction des demandes et tarification

Démarche et instruction des dossiers

1. Remplir un formulaire en ligne (possibilité de soumettre la demande en suédois, finlandais et en anglais)
2. Soumission d'un plan d'utilisation des données par le porteur de projet
3. **Facturation**
4. Findata évalue la correspondance de la demande aux préconisations de l'Act on the Secondary Use of Health Data (55/2019)
5. Findata peut éventuellement demander une approbation du Conseil Éthique du National Institute for Health and Welfare (THL).
6. Une fois la demande approuvée, Findata enverra les demandes de prélèvement d'échantillons et les métadonnées aux responsables du traitement des données.
7. À la réception des données pertinentes (non anonymisées) des responsables de données, Findata procédera au couplage (en utilisant le numéro d'identification personnel national attribué aux individus à la naissance et utilisé par tous les responsables du traitement) et à la pseudonymisation.
8. L'accès des utilisateurs finaux aux données individuelles est autorisé uniquement par le biais d'un environnement de recherche sécurisé pendant une période définie (l'accès aux données expire). En plus d'être l'autorité chargée de délivrer les permis de données, Findata est également le fournisseur des données à l'utilisateur final.

Exigences préalables et modèle économique

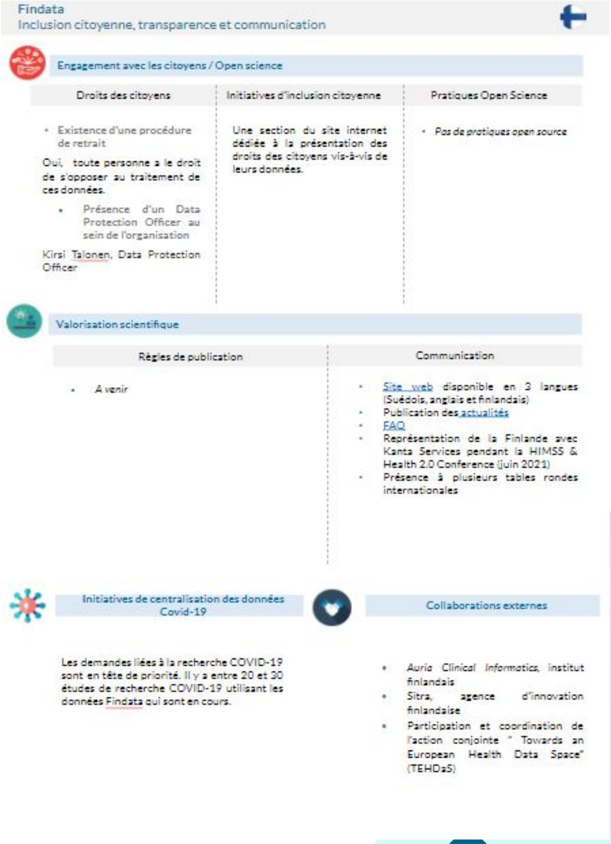
Exigences préalables	Modèle économique
Être chercheur affilié à une institution académique, finlandais ou européen Fournir descriptions du contenu des statistiques ou des données à extraire (registres) Spécification d'un objectif conforme aux dispositions de la loi Fournir des détails pour la date de livraison du matériel de données Pas d'obligation de figurer dans une liste publique des projets Pas d'obligation de description en langue simple pour le public	Tarification de l'examen des demandes et de la délivrance des permis: → Tarification de l'examen des demandes: 1000 euros pour tous les acteurs → Délivrance du permis dépend du porteur de projet (différenciation entre entreprises finlandaises et européennes et autres entreprises étrangères, et chercheurs: 1000, 3000 et 500 euros) Prix demandé par les RDD pour l'extraction des données Tarification de l'accès au remote access environment

Instruction des dossiers

- ✓ Formulaire de demande d'accès, y compris:
- ✓ Description du projet de recherche
- ✓ Description de la base juridique du projet
- ✓ Description des chercheurs
- ✓ Description des collaborateurs
- ✓ Description des données requis (quelles bases et quels criteres de selection des individus)
- ✓ Date d'achèvement prévue

Processus de prise de décision	Géré par Findata Management Group
Formalisme de l'avis	*Non communiqué*

8



Findata
Inclusion citoyenne, transparence et communication

Engagement avec les citoyens / Open science

Droits des citoyens	Initiatives d'inclusion citoyenne	Pratiques Open Science
Existence d'une procédure de retrait Oui, toute personne a le droit de s'opposer au traitement de ces données. Présence d'un Data Protection Officer au sein de l'organisation Kirsi Talonen, Data Protection Officer	Une section du site internet dédiée à la présentation des droits des citoyens vis-à-vis de leurs données.	*Pas de pratiques open source*

Valorisation scientifique

Règles de publication	Communication
A venir	Site web disponible en 3 langues (Suédois, anglais et finlandais) Publication des actualités FAQ Représentation de la Finlande avec Kanta Services pendant la HIMSS & Health 2.0 Conference (juin 2021) Présence à plusieurs tables rondes internationales

Initiatives de centralisation des données Covid-19

Les demandes liées à la recherche COVID-19 sont en tête de priorité. Il y a entre 20 et 30 études de recherche COVID-19 utilisant les données Findata qui sont en cours.

Collaborations externes

- *Auria Clinical Informatics*, institut finlandais
- Sitra, agence d'innovation finlandaise
- Participation et coordination de l'action conjointe " Towards an European Health Data Space" (TEHDaS)

Overview of analyzed structures

Category 1:

Structures **hosting and granting access** to health data

- 🇩🇪 *DaTrav (implementation in process)*
- 🇧🇪 Healthdata.be
- 🇰🇷 HIRA
- 🇩🇰 Danish Health Data Authority
- 🇺🇸 Regenstrief
CMS
- 🇫🇮 Findata
- 🇳🇴 *Health Data Program (implementation in process)*
- 🇬🇧 UK Biobank
CPRD
SAIL Databank

Category 2:

Structures impacting the **governance and quality of health data**

- 🇩🇪 Medical Informatics Initiative
- 🇨🇦 Health Data Research Network
- 🇪🇪 X-Road and Estonian e-Health system
- 🇺🇸 National Library of Medicine
- 🇮🇱 Clalit Research Institute
- 🇬🇧 HDR-UK

1/2 Description of Category 1 Structures: Data Access Platforms

CMS *created in 1965*	Centers for Medicare and Medicaid Services (CMS) was established in 1965 and hosts the data collected in the framework of the Medicaid and Medicare programs. CMS provides access to the generated data to researchers and for-profit organizations through the Office of Enterprise Data and Analytics (OEDA); which manages the Chronic Conditions Data Warehouse, the Research Data Distribution Center, and the CMS Research Data Enclave.
CPRD *created in 2011*	Clinical Practice Research Datalink (CPRD) is a research service within the UK Department of Health. CPRD data is derived from a UK network of more than 2,000 primary care practices and includes 50 million patients of which 16 million are currently registered active patients with at least 20 years follow-up for 25% of patients. CPRD provides access to academic and private sector researchers from around the world to support observational public health research.
Danish Health Data Authority *created in 2015*	In the absence of a centralized program, the Danish Health Data Authority which is part of the Ministry of Health handles many public health registries and provides a research support service (Forskerservice) for researchers health data applications to national and regional authorities as well as a national data map.
DaTrav *planned for 2022*	DaTraV, the German Research Data Center, is the future data platform for research (healthcare claims data of insured individuals in Germany). Soon it will also be a hub for electronic medical records data. Data will be provided by the German Federal Association of Health Insurance Funds (GKV-SV).
Findata *created in 2019*	Findata is the Finnish authority that issues permits for access to health and social data in Finland. It is a one-stop shop for a broad range of data sources. Since its operation in January 2020, the authority has received 411 data access requests and has issued 265 decisions (as of end of April 2021). Its mandate is defined by the 2019 Act of Secondary Use of Health Data.
Healthdata.be *created in 2014*	Healthdata.be, managed by Sciensano, gathers individual health data collected by national and regional administrations. The platform is responsible for all data flows related to research within the national e-health system. Healthdata.be offers the following services: a) the provision of pseudonymized data from the Data Warehouse and b) the operationalization of a new collection of pseudonymized data from healthcare actors.

2/2 Description of Category 1 Structures: Data Access Platforms

🇰🇷	**HIRA** *created in 2000*	HIRA (Health Insurance Review and Assessment) is a public institution responsible for reviewing health insurance claims in South Korea. HIRA's database hosts all National Health Insurance Service data and can be accessed by researchers since 2015 via HIRA Data Services.
🇳🇴	**Norwegian Health Data Program** *planned for 2022*	The Health Data Program (HDP) is developed by the Norwegian Directorate of eHealth, established in 2016, to increase the efficiency of access to health data. The programme provides a one-stop shop for Norwegian health data, accelerating and facilitating current procedures, and also enables the deployment of the Health Data Analysis Platform, a data analysis environment for researchers.
🇺🇸	**Regenstrief Institute** *created in 1967*	Regenstrief Institute is a non-profit organization and an international leader in the field of electronic medical records and data integration; handling state-level data from the Indiana Network for Patient Care (INPC) and other specialized data sources. The mission of Regenstrief Data Services is to leverage the data and IT resources of the Regenstrief Institute to support the search for innovative healthcare solutions
🇬🇧	**SAIL Databank** *created in 2007*	The SAIL system is a national architecture for e-health research and evaluation, availing pseudonymized individual-level data for research purposes. The data covers the Welsh population. Researchers can access a wide range of data collected regularly over a period of up to 20 years. SAIL data is used in many areas of research to assess the impact of health policies.
🇬🇧	**UK Biobank** *created in 2006*	UK Biobank is a cohort of 502,536 participants between the ages of 40 and 69. The data collected, since 2006, are derived from questionnaires sent to the participants and from biological data and samples collected. UK Biobank grants access to these data to researchers from the public and private sectors for research of public interest, after registering on the UK Biobank database (Access Management System).

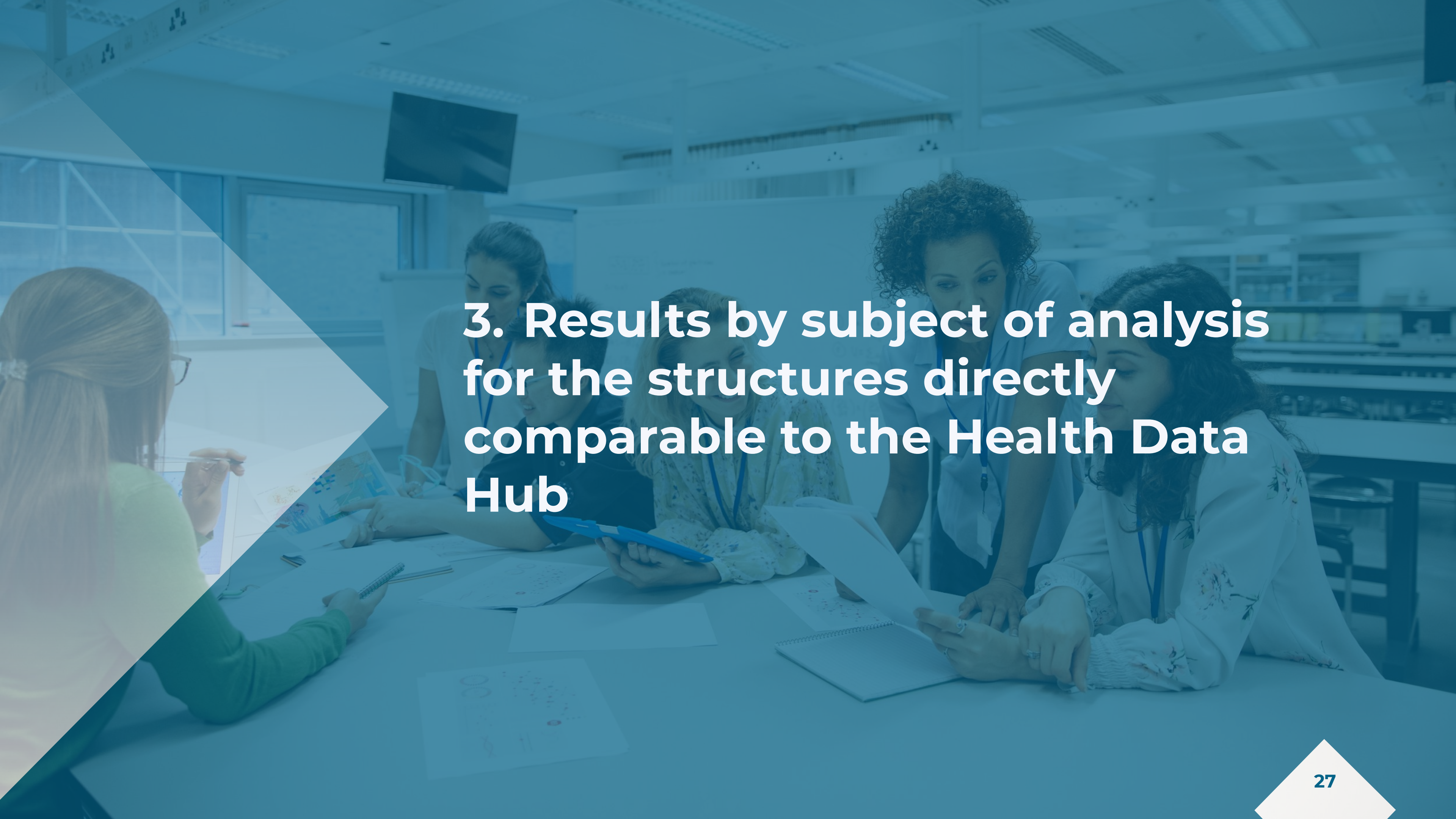
Description of Category 2 Structures : Structures influencing Health Data Quality and Governance

Clalit Research Institute *created in 2014*	Clalit Research Institute is the research arm of Clalit Health Services; one of the largest healthcare providers in Israel. The institute works on health innovation, and utilizes data from the large Clalit database.
HDR-UK *created in 2019*	Health Data Research UK (HDR-UK) is the UK's health data institute; which aims at bringing together nearly 86 health data organisations (within the UK Health Data Research Alliance) in 56 offices across 32 locations. The path to data access is offered via the Innovation Gateway set up by the Institute; which also offers a visualisation of the metadata catalogues of partner structures.
Health Data Research Network (HDRN) *created in 2020*	HDRN is a not-for-profit network created with the mission to improve health and well-being by making data accessible to researchers, institutions and government agencies across Canada for research that will foster improved health outcomes for all Canadians. HRDN facilitates data access procedures for researchers, institutions and government agencies. One of the goals is to explain extremely complex and heterogeneous multi-jurisdictional (provincial, territorial, and pan-Canadian) data access processes within the healthcare sector across Canada.
Medical Informatics Initiative (MII) *created in 2018*	The MII is established between 4 German university research consortia for the exchange and use of health data for medical research. The sites have agreed on a model for data sharing and exchange, including uniform access procedures and transfer points at all participating sites, as well as a harmonised consent form. Several working groups are in place, for example on citizen involvement and consent for secondary use of their data.
National Library of Medicine *created in 1836*	The National Library of Medicine (NLM) carries out its mission of enabling biomedical research, supporting health care and public health. Created in 1836, the NLM is part of the U.S. Department of Health and Human Services. It is the world's largest biomedical library; producing electronic information resources on a wide range of key topics for the healthcare system.
X-Road *created in 2001*	The Estonian X-Road® infrastructure, based on the X-tee cloud, enables the exchange and interaction of data between the different information systems of the country's public and private sector e-services. This system provides a comprehensive overview of the data flows of the entire ecosystem for the government.

Two additional focus points

Focus on Findata:
The study conducted reveals in particular that the HDH is very close to the Finnish model, Findata. The example will be presented in detail.

Focus on Data altruism
The European Commission emphasises that data altruism or 'data solidarity', a principle by which organisations or/and citizens can decide to make their data available for secondary use purposes, will have a more prominent role at the European Union level in the future. Some modalities and frequency of data altruism initiatives across the EU will be detailed.



3. Results by subject of analysis for the structures directly comparable to the Health Data Hub

Focus : Findata

Findata is the Finnish health data permit authority. It is a true **one-stop shop for health and social data in Finland,** operational since January 2020.

Description

The objective of Findata is to provide fast, secure and easy access to individual-level (social and health) data .

Findata grants access to data from multiple registries and data controllers in Finland, collects the requested data from these registries and data controllers to combine them, pseudonymizes them, and finally delivers the data via a secure remote project environment.

Legal basis

The Act on the Secondary Use of Health and Social Data (552/2019) is the legal basis for the implementation of Findata. This act specifies the purposes for which access to data may be requested: statistical purposes, scientific research, for innovation and development, education, supervision and guidance of social and health authorities.

Governance

The Steering Group was charged with defining the provisions of the *Act on the Secondary Use of Health and Social Data.* It includes representatives of data controllers, and is responsible for guiding and developing Findata's operations.

Findata is supervised by Valvira (a national agency of the Finnish Ministry of Health), the National Institute of Health, and the Data Protection Agency.

Data sources

Findata grants permissions for data collected in public and private sector services that are part of the national data sources defined by law, including:

- Finnish Institute of Health and Welfare
- National Health Insurance Fund
- Statistics Finland
- Digital and Demographic Data Services Agency
- University Hospital Helsinki
- Finnish Cancer Registry
- Other university hospitals
- Finnish Centre for Pensions
- Cities of Helsinki, Espoo, Vantaa…

Analysis of the results : Platforms

Observed common trend

Benchmarked platforms (category 1) guarantee an **operational technological platform for data storage**, developed on a national or non-European scale, guaranteeing **high standards of quality and security** - particularly for the **protection of the privacy and identifiable data**.

NB: The future German platform DaTraV is not included.

Technological platform

Common practice

6 out of the 10 category 1 structures **have a non-European technological platform**

CPRD, UK Biobank, HDP, SAIL Databank, healthdata.be, Danish Health Data Authority

Common practice

4 out of the 10 category 1 structures **have a home-grown technological platform**

Findata, HIRA, CMS, Regenstrief Institute

Highlight

Findata

Findata platform is based on the Finnish **ePouta** infrastructure for sensitive data: the cloud offers a **complete service including ePouta, ePouta Remote Desktop and a sensitive data management platform**.

How to access data

Common practice

6 out of 10 of the category 1 structures offer **data access via a secure project space**

Danish Health Data Authority, Findata, healthdata.be, HDP, CMS, SAIL Databank

Common practice

3 out of 10 of the category 1 structures offer **direct data transfer to the user**

CPRD, UK Biobank, HIRA

Highlight

Regenstrief Institute

Double system

Anonymised data is transferred directly to the user, but personal data can be accessed in the Regenstrief secure research environment

Analysis of the results : Business model

Observed common trend

Fees for granting access to data, for data extraction services, or for using the secure project environment are common. Two practices are observed: **differentiation of fees among different actors and guaranteeing fixed fees for all actors accessing the data.** *(HDP is not included)*

Fees to access data

Common practice

7 out of 10 Category 1 platforms charge for **project-specific services**
Findata, Danish Health Data Authority, healthdata.be, SAIL Databank, CMS, HIRA, DaTraV

Ad hoc practice

4 out of 10 category 1 platforms charge for **data licenses/permits**
Findata, Regenstrief Institute, CPRD, UK Biobank

Highlight

Findata
Applies a fee for access review, permitting, remote access to the project environment (per month), services associated with its preparation (per hour), and consults with Data controllers for data extraction fees.
Applies a differentiated pricing according to the type of actors

Fee categories

Common practice

6 out of 10 category 1 platforms apply different pricing to **actors according to their type (national/international companies, national/international researchers)**
CMS, Findata, CPRD, Healthdata.be, Regenstrief Institute, DaTraV

Ad hoc practice

4 out of 10 category 1 platforms **apply the same pricing without differentiation**
Danish Health Data Authority, HIRA , UK Biobank, SAIL Databank

Highlight

Danish Health Data Authority
Applies pricing to all actors in the same way.
On the other hand, researchers and public institutions are eligible for assistance from the Danish Research Organisation (KOR) to finance the platform's services, for which industrial entities are not eligible

Analysis of the results : Scientific valorization

Observed common trend

Publication rules are not homogeneous. However, all the platforms studied require to be mentioned in publications.

Open science practices can take the form of publishing research results or availing relevant documentation to the public.

Findata, HDP, and DaTrav have not yet set publication requirements.

Publication requirements

Common practice

Out of 8 category 1 structures that provided the information, all of them request to **be mentioned in the publications**.

Ad hoc practice

2 out of 8 of the Category 1 structures that provided the information **may grant a period of exclusivity (9 to 12 months).**
Healthdata.be and UK Biobank

Highlight

CPRD

An audit committee is responsible, on an ad hoc basis, for verifying the concordance between the research protocol submitted at the time of the data access request and the publications.

Open science

Common practices

Out of 10 Category 1 structures that provided information on this topic (including Findata), 6 requested that **publications be published and made freely available**.
UK Biobank, CMS, CPRD, SAIL Databank, Danish Health Data Authority, healthdata.be

7 out of 10 structures that provided information on this topic have **online documentation (data guide, metadata catalogues)**
HIRA, healthdata.be, UK Biobank, SAIL Databank, CPRD, Danish Health Data Authority, Regenstrief

Highlight

HIRA

Statistics on health services (health care expenditures and utilization - including prescriptions, medical conditions and providers) are publicly available through the HIRA open data health care system (http://opendata.hira.or.kr).

Results by Subject of Analysis: Public Relations/ Communication/ Advertising

Observed common trend

The structures analyzed have **diversified communication and public relations strategies**, some more advanced than others, but often with the common goal of broadening their footprint on social networks by targeting citizens to **promote information on health data and to immerse the structure in the virtual ecosystem through the creation of an easily accessible website.**

Communication practices

Common practice

14 out of 17 of all the benchmarked structures **have an operational website and publish their news, details of their access procedure if applicable,** and documentation targeting applicants and the general public.

Ad hoc practices

5 structures out of 17 (HIRA, UK Biobank, HDR-UK, CMS and MII) **publicly organize online events**

3 structures out of 17 (**Regenstrief, CMS and HDR-UK) have their own online podcast** ("The Problem" ,"CMS beyond the policy" , "The Genetics Podcast")

Highlights

HDR-UK

HDR-UK has elaborated a real communication strategy, a network of nearly 60 people has been mobilized to manage the website as well as the different social networks.

SAIL Databank

Publication of comments on the SAIL Databank experience from the project coordinators on the homepage of their website.

Analysis of the results : Citizen engagement

Observed common trend
Information on **citizens' rights** (consent and existence of a withdrawal procedure) is unanimously made available; their **inclusion in the governance of the structures** is a more ad hoc practice. *HDP and DaTrav did not provide information.*

Citizen's rights

Common practice

5 Category 1 structures **allow citizens to withdraw their data**
CPRD, UK Biobank, Findata, SAIL, Regenstrief

Ad hoc practice

2 Category 1 structures are **entirely based on citizen data donation (opt-in)**
CPRD and UK Biobank

Highlight

Danish Health Data Authority
The structure guarantees ***right-to-redress***: patients can request correction of their information.
Availability of a **Vævsanvendelsesregisteret** database; hosting citizens' requests for secondary use of their health data.

Information for citizens

Common practices

All structures offer an open website where categories **"citizens' rights"** or **"privacy"** are available.

Ad hoc practice

2 category 1 structures propose a **"self-access"**: citizens can **obtain a copy and access the data concerning them.**
Regenstrief Institute and the Danish Health Data Authority

Highlight

Understanding Patient Data
An initiative of the Wellcome Trust, Medical Research Council and Public Health England to support the understanding of secondary use of patient data.

Citizens represented in structure governance

Ad hoc practice

3 Category 1 structures include **citizens in their governance**
SAIL Databank, Healthdata.be and Regenstrief Institute

Highlight

SAIL Databank

The SAIL Public Engagement Team organizes events to promote SAIL's activities to the public. Data perception surveys with the SAIL Databank **Consumer Panel** are also organized.

Analysis of the results : Game changers in the COVID-19 crisis

Observed common trend

Across all categories combined, the structures have mobilized resources to **support the centralization and exploitation of COVID-19 data,** through key initiatives and data visualization dashboards, and to **enable project coordinators to access data quickly**.

COVID-19 Data Initiatives

Common practice

All Category 1 data platforms (*except HDP and DaTraV*) had put in place **accelerated data access procedures for COVID-19 research**.

Ad hoc practice

8 structures out of 17 offer a **metadata catalog or other tools for visualizing COVID-19 data.**
Regenstrief Institute, Healthdata.be, SAIL Databank, UK Biobank, CPRD, HDRN, CMS, HDR-UK

Highlights

HDR-UK

International COVID-19 Data Alliance (ICODA) is an international platform that allows researchers to access global data to quickly obtain information on COVID-19 and thereby accelerate research.
The alliance includes HDRN, UK Biobank and CPRD.

Medical Informatics Initiative

CoCos Initiative: a multi-stakeholder initiative to establish uniform COVID-19 data formats.
German Corona Consensus: Prioritizes data collection from university hospitals for a compact research database.

Focus : Data altruism

The European Commission points out the importance of developing, at national and European levels, the concept of data altruism. This would allow organizations, and even citizens, to make data available for research purposes.
Only **two EU member states have so far implemented a plan to introduce a national data altruism system:** Denmark and Germany.
Some initiatives comparable to altruistic practice have also been observed in the United Kingdom.

Germany

Starting in 2023, the **Patient Data Protection Act** will give insured citizens, the ability to make their electronic medical record data available to researchers.

Denmark

A 2-year strategy for the implementation of altruistic practices has been developed.
Secure spaces for hosting the data of citizens who wish to share it will be opened from the **Sundhed.dk** platform.

So-called "bottom-up" non-governmental initiatives are also emerging, such as **the NEXT database** to facilitate the registration of citizens who wish to participate in clinical trials.

United Kingdom

Several British initiatives have been identified.

The **SHARE (Scottish Health Research Register)** initiative of the National Health Service in Scotland is a registry of patients over 11 years of age who consent to participate in research based on their health data.

According to the EUHealthSupport Consortium study, 14 countries are willing to implement a model of data altruism at the national level, and 11 countries are willing to implement it at the European level.

Source: Assessment of the EU Member States' rules on health data in the light of GDPR
Link: https://ec.europa.eu/health/sites/health/files/ehealth/docs/ms_rules_health-data_en.pdf

4. Best practices from other actors of the international health data ecosystem

Going further: international actors in the health data ecosystem

Data governance structures with slightly distinct mandates than the HDH have been identified and provide insights of best practices in for the secondary use of health data

- Federate the actors in the health data ecosystem
- Provide evidence (health data) to implement new public policy (evidence-based policy-making
- Support project coordinators
- Intensify and encourage data sharing between the research community and the health system
- Establish good practices and quality standards
- Provide data infrastructure to connect governmental services

Going further: international actors in the health data ecosystem

Federate actors of the health data ecosystem

HDR-UK

The **Innovation Gateway** is a unique portal for researchers, enabling visualization of available data through an enriched catalog of **metadata from members of the UK Health Data Research Alliance**.

This Gateway includes an overview of the GPES Data for Pandemic Planning and Research COVID-19 datasets from the NHS Digital.

HDR-UK also includes **8 research and data collection hubs across the** UK: BREATHE, DATA-CAN, Gut Reaction, INSIGHT, NHS DigiTrials, PIONEER, DISCOVER-Now, BHS Data Science Centre.

Provide evidence (health data) to inform and improve public policy (evidence-based policy-making)

Medical Informatics Initiative

Data Integration Centres

Medical Informatics Initiative four consortia have created **Data Integration Centers (DICs) within their university hospital sites**.

DICs create the technical and organizational conditions necessary for data sharing across multiple sites.

Support project coordinators

Health Data Research Network

The **Strategy for Patient-Oriented Research Canadian Data Platform (SPOR CDP),** aims to create a system to support access to health data.

This service includes the **Data Access Support Hub (DASH)**, a single-window data access service portal for researchers requiring multi-jurisdictional data in Canada.

Going beyond international actors in the health data ecosystem

Intensify and encourage data sharing between the research community and the health system

National Library of Medicine

NLM develops and distributes the NLM classification system. This system is widely used by academic medical libraries to organize their collections.
NLM staff **work with international cataloguing and archival standards**, including Dublin Core, MARC and other ISO standards, as well as international data standardization and terminologies.

Establish good practices and quality standards

Clalit Research Institute

The mission of the Clalit Health Services research institute is to use its large database, mainly enriched with electronic medical records, to **develop predictive health tools** and to develop new innovative health practices. The organization offers multiple use cases published in major scientific journals such as Nature or BMJ.

Provide data infrastructure to connect governmental services

X-Road

X-Road is **a registry and data exchange system between different public services,** based on the **X-Tee cloud.** This infrastructure drastically facilitates exchanges between the government and citizens. The government and healthcare actors can directly access certain information on the database, and citizens can access all services, 99% digitized, thanks to a unique identifier (eID).

HEALTH
DATA HUB

Follow us on social media!