HEALTH DATA HUB

Building a data-driven healthcare system: Strategies, and insights from the Health Data Hub

by Emmanuel Bacry
September 12th, Paris

1

1. Introduction | The Health Data Hub, a public health data service

The uses of health data are increasing and it has become essential to ensure access to data sources as quickly as possible. Created at the end of 2019, the Health Data Hub is a public body tasked with facilitating access to health data for projects in the public interest, following the granting of open access to the French National Health Data System (Système national des données de santé – SNDS) in 2016.

A **single gateway** for health data in France

A **secure,** state-of-the-art **platform**

A **data catalogue** including one of the world's largest healthcare claims databases

The HDH actively supports 213 innovative projects to this date.

HEALTH DATA HUB

After five years, where do we stand?

Production launch of a technology platform in just one year

Recruitment of staff to reach 116 people by 1 January 2025

Support for 213 projects, a third of which are led by industrial actors

Major European involvement, notably as leader of the HealthData@EU Pilot project, to **build the first version of the EHDS**

More than 100 active partnerships, resulting in the development of links with the entire ecosystem, including civil society
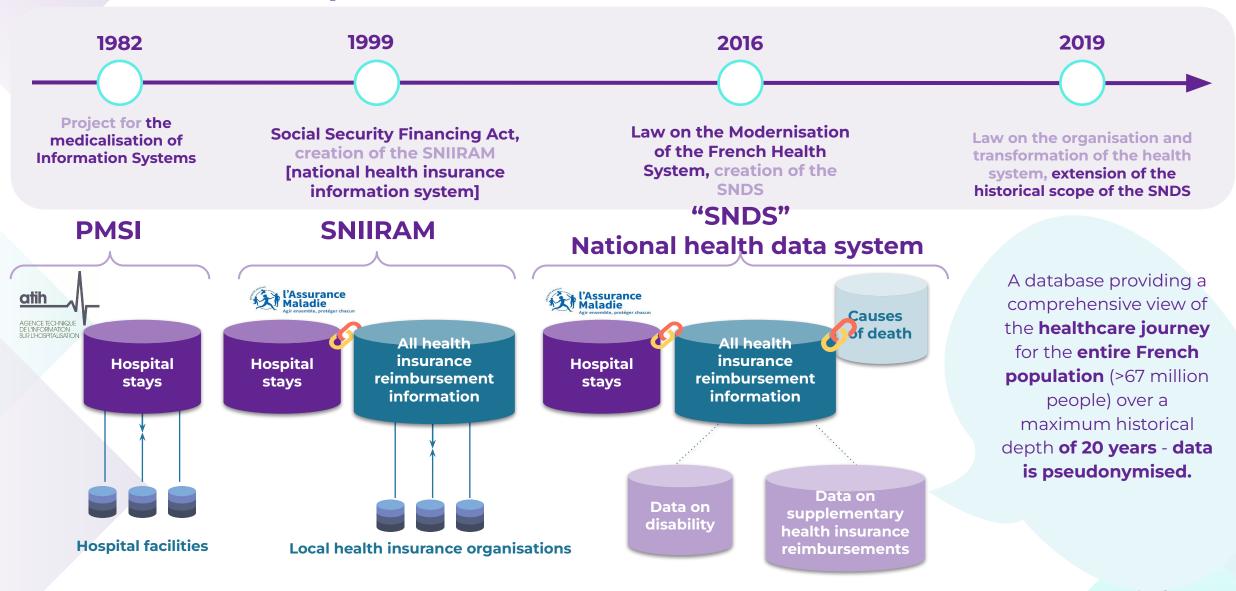
A major impetus for the constitution of new databases: AAP on health data warehouses (€75M), P4DP project (€10M)

More than 650 public talks since 2021 and **participation in major projects such as PariSanté Campus, FIAC, DARWIN**

213 **projects supported by the HDH**, of which 155 use the technology platform

48 projects in the regulatory phase (data access request)

36 projects in the data preparation phase

42 projects running on the platform

29 projects completed, or with intermediary results

24 **Calls for project**
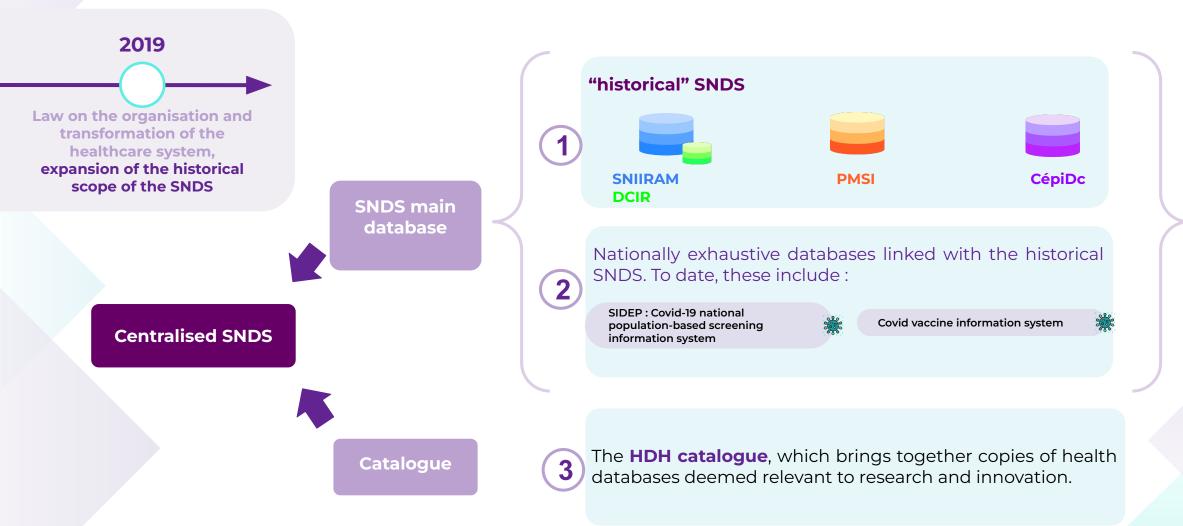
What is the expanded SNDS ? Context of creation

1982

Project for **the medicalisation of Information Systems**

1999

Social Security Financing Act, creation of the SNIIRAM **[national health insurance information system]**

2016

Law on the Modernisation of the French Health System, creation of the SNDS

2019

Law on the organisation and transformation of the health system, **extension of the historical scope of the SNDS**

PMSI

atih
AGENCE TECHNIQUE DE L'INFORMATION SUR L'HOSPITALISATION

Hospital stays

Hospital facilities

SNIIRAM

l'Assurance Maladie
Agir ensemble, protéger chacun

Hospital stays

All health insurance reimbursement information

Local health insurance organisations

"SNDS" National health data system

l'Assurance Maladie
Agir ensemble, protéger chacun

Hospital stays

All health insurance reimbursement information

Causes of death

Data on disability

Data on supplementary health insurance reimbursements

A database providing a comprehensive view of the **healthcare journey** for the **entire French population** (>67 million people) over a maximum historical depth **of 20 years** - **data is pseudonymised.**

Following the law of 2019, the scope of the SNDS has been expanded with the HDH catalogue

2019

Law on the organisation and transformation of the healthcare system, **expansion of the historical scope of the SNDS**

SNDS main database

Centralised SNDS

Catalogue

"historical" SNDS

1

SNIIRAM DCIR

PMSI

CépiDc

2

Nationally exhaustive databases linked with the historical SNDS. To date, these include :

SIDEP : Covid-19 national population-based screening information system

Covid vaccine information system

3

The **HDH catalogue**, which brings together copies of health databases deemed relevant to research and innovation.

The data catalog will soon be expanded to include other databases addressing a wide range of public health issues.

The **SNDS catalog** is a collection of databases that is not fixed in stone, but is built up iteratively. The decree of May 12, 2022 lists the **first 10 databases** to be included in the catalog, to which will be added the databases identified to join the catalog.

The integration of 12 other databases is currently being examined by the CNIL:

Base e-sis: breast cancer

Base ATU: early access

BNDMR: rare diseases

EPICOV: Covid-19

ESME: breast cancer

E-must: myocardial infarction

HEPATHER: hepatitis B or C

MDO: list of mandatory diseases

Memento: Alzheimer disease

OSCOUR: emergency monitoring

Databases from call for project

ARAC: remaining expenses

APSOREN, TARPON: emergency medicine, traumatology

SEDAAR: ophthalmology

APRIORICS, PRECISION PREDICT, NetSARC: oncology

Databases supported in their construction

P4DP: creation of a data warehouse for community care

ISIS, REALIGIST: creation of an oncology reference database
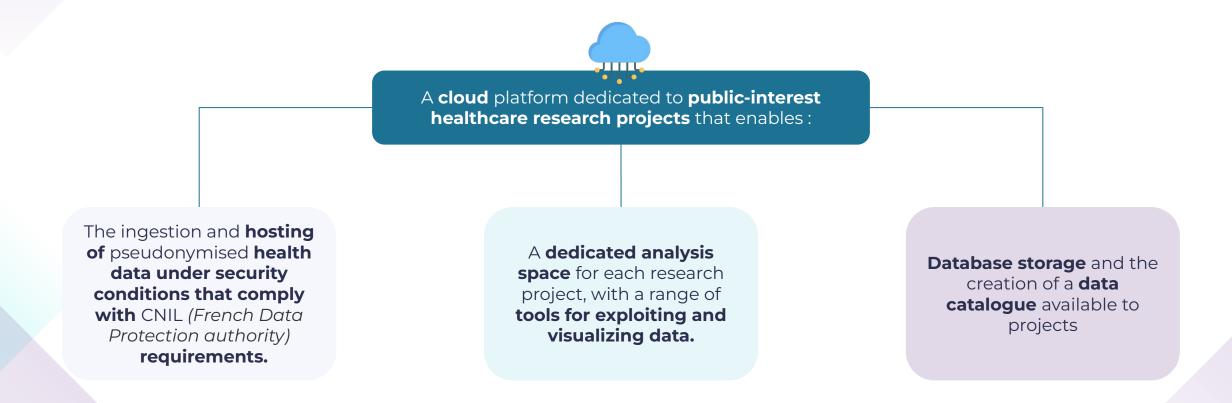
Other relevant databases

INTEGRA: hospital cohort on metabolic diseases

UroCCR, biological clinical registry for kidney cancer

SNDS
Système national des données de santé

HEALTH DATA HUB

What is the technological platform? | Focus

A **cloud** platform dedicated to **public-interest healthcare research projects** that enables :

The ingestion and **hosting of** pseudonymised **health data under security conditions that comply with** CNIL *(French Data Protection authority)* **requirements.**

A **dedicated analysis space** for each research project, with a range of **tools for exploiting and visualizing data.**

Database storage and the creation of a **data catalogue** available to projects

It is part of the Health Data Hub's mission to simplify **access to healthcare data, in a state-of-the-art technological and IT secure** environment.

3. Technological offer | current offer*

	Programming languages and IDEs	Big Data infrastructures	Data visualisation	Data storage	Images	Collaboration
Available tools	python, R, jupyterlab, RStudio	Apache Spark	Superset	MySQL, ORTHANC, PostgreSQL	cytomine.ORG, OHIF Viewer	GitLab, Gitea, git
Use	• Data exploration • Machine learning • Image processing	• Queries and processing of a high volume of data	• Visualisation of data	• Structuring of data • Queries	• Visualisation and annotation of images	• Code versioning • Development follow up

Complexity to take into account if a project leader wishes to use a tool outside the current technological offer

How is the Health Data Hub being used in practice?

The Health Data Hub is the only **French cloud technology platform**

- that provides health data projects with **access to elastic storage and high computing capacity**.

- that allows **different data sources to be linked**

- **designed for use by a large number of users**

- while meeting the **highest security standards**

Supporting healthcare professionals in an increasingly complex clinical environment, with use cases such as the characterization of alerts for drug interactions, based on the main SNDS database

Saving time, improving screening, and reducing diagnostic delays — notably through the development of a diagnostic support tool for 10,000 MRIs

Develop early warning systems for patient follow-up, including the development of a warning tool for healthcare professionals based on data from connected medical devices, annotated with medico-administrative data

Offer patients the best long-term treatments, with use cases comparing care paths using cohort data enriched with medico-administrative data when clinical trials are not possible

The Health Data Hub works with the Government to prepare a national doctrine for hospital data warehouses

Announce of **laureates** in two waves : April 2023 & 2024 T1
Call closed in April 15, 2023

In 2022, the French government has launch a call for project to support the definition of a **national vision** for the creation of a national network of **hospital health data warehouses**.

Role of the Health Data Hub

1 Provide operational support for the interministerial committee in organising the call for projects
2 Support the laureates in the implementation of their project
3 Manage the community

75M€ allocated to laureates of the call

50% of expenditure covered

3 research and innovation projects to be launched min.

12 to 40 months

Creating a National doctrine for hospital data warehouses will allow to :

- **Standardise practices and tools at the national level**

- **Define a target territorial coverage**
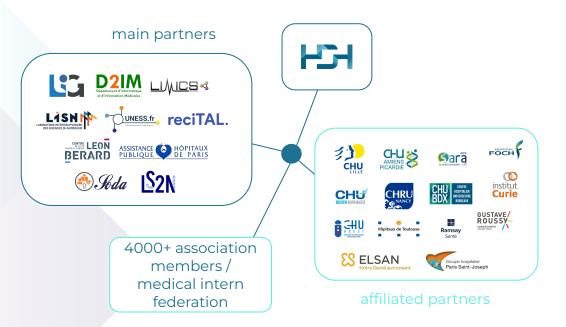
- **Build a sustainable financial model**

FINALISED

Create a Common data foundation for hospital data warehouses

Healthcare data access governance

Common database standardization and interoperability

WORK IN PROGRESS

Data catalogue

Sustainable financing of health databases and fees for the provision of healthcare data

Standard data provision contract

The HDH commitment to support AI | "PARTAGES" a candidate consortium for the "Digital Commons" call, to create medical LLM in French and a corpus of synthetic medical records

The HDH has created a consortium to respond to the "digital commons for generative AI" call for projects, launched as part of France 2030. The consortium is mobilising cutting-edge skills to **produce impactful deliverables that will be 90% open source and reusable by the entire medical, academic and industrial scientific community**.

main partners

affiliated partners

4000+ association members / medical intern federation

Resulting in :

- **A basic medical LLM in French,** *open source*

- **A training database of 10,000 fictional medical reports**

- **Specific algorithms corresponding to initial use cases or training** *relevant to the ecosystem*

- **A federated national platform for validating models** *on real hospital data that can be reused as part of the Grand Challenges program*
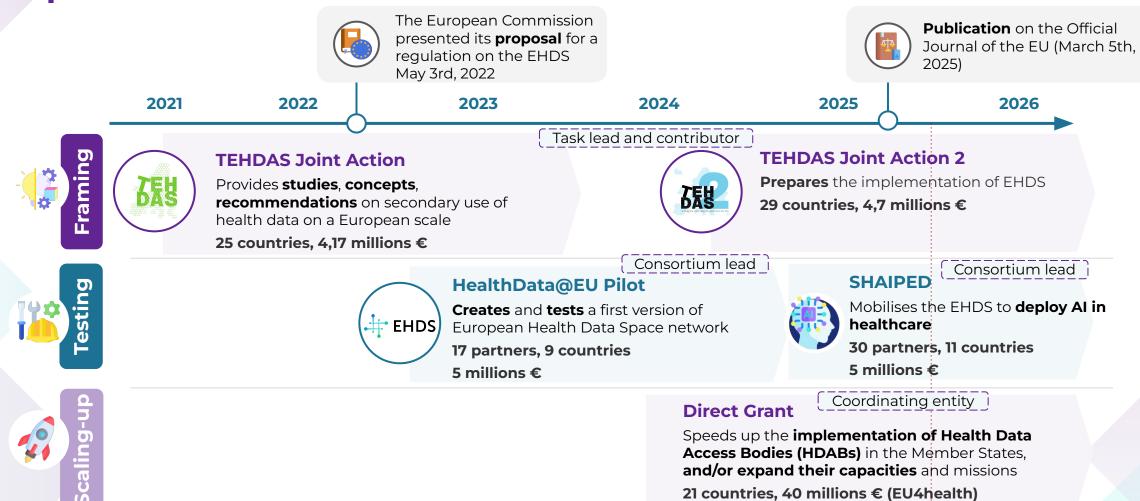
- **Methodological guides for algorithm refinement** *(fine-tuning) by hospital staff using their own data*

Growing initiatives for European Health Data Space implementation

The European Commission presented its **proposal** for a regulation on the EHDS May 3rd, 2022

Publication on the Official Journal of the EU (March 5th, 2025)

2021 **2022** **2023** **2024** **2025** **2026**

Framing

Task lead and contributor

TEHDAS Joint Action

Provides **studies**, **concepts**, **recommendations** on secondary use of health data on a European scale

25 countries, 4,17 millions €

TEHDAS Joint Action 2

Prepares the implementation of EHDS

29 countries, 4,7 millions €

Testing

Consortium lead

HealthData@EU Pilot

Creates and **tests** a first version of European Health Data Space network

17 partners, 9 countries

5 millions €

Consortium lead

SHAIPED

Mobilises the EHDS to **deploy AI in healthcare**

30 partners, 11 countries

5 millions €

Scaling-up

Coordinating entity

Direct Grant

Speeds up the **implementation of Health Data Access Bodies (HDABs)** in the Member States, **and/or expand their capacities** and missions

21 countries, 40 millions € (EU4health)

HEALTH DATA HUB

Overview of the SHAIPED project

Supporting **H**ealth Data Access Bodies to establish **AI** **P**athways **E**nabling **D**eployment of AI as medical device tools

4M€
in EU funding

3 years duration

30
Partners throughout Europe

11
Member States involved

Research Institutions / Hospitals / Public Instances / Start-ups

11 HDABs and other related organizations

Within the 3 next years, SHAIPED is expected to provide the foundations and first results of an European AI space for health

SHAIPED aims at identifying, building new and strengthening existing pathways for AI medical device development, testing and deployment provided by Health Data Access Bodies

Framework Design

Framework Testing

Pathways
More effective pathways for developing, testing and deploying AI medical devices leveraging EHDS and HDABs

Capabilities
Creation and implementation of the **necessary capabilities for HDABs** and their ecosystem (innovation, regulatory and data) to support these pathways

Recommendations
Guidelines, recommendations and **regulatory** backdrop to implement those pathways

Concrete outcomes
Testing of established pathways (3 use cases) with **outcomes**, **opportunities** and **challenges** to consider

Do you have any questions ?

HEALTH DATA HUB

Follow us on social media!