



NHIS GLOBAL FORUM **2025**

**The French Health Data Hub:
towards data-driven
healthcare**

Dr Salam ABBARA

**The French Health
Data Hub:
towards data-driven
healthcare**

Dr Salam ABBARA

***NHIS Global Forum 2025, Seoul
11-12 June 2025***

Agenda

I. **An introduction to the activities of the Health Data Hub**

Context of creation and main missions

I. **Secondary health data use landscape in France**

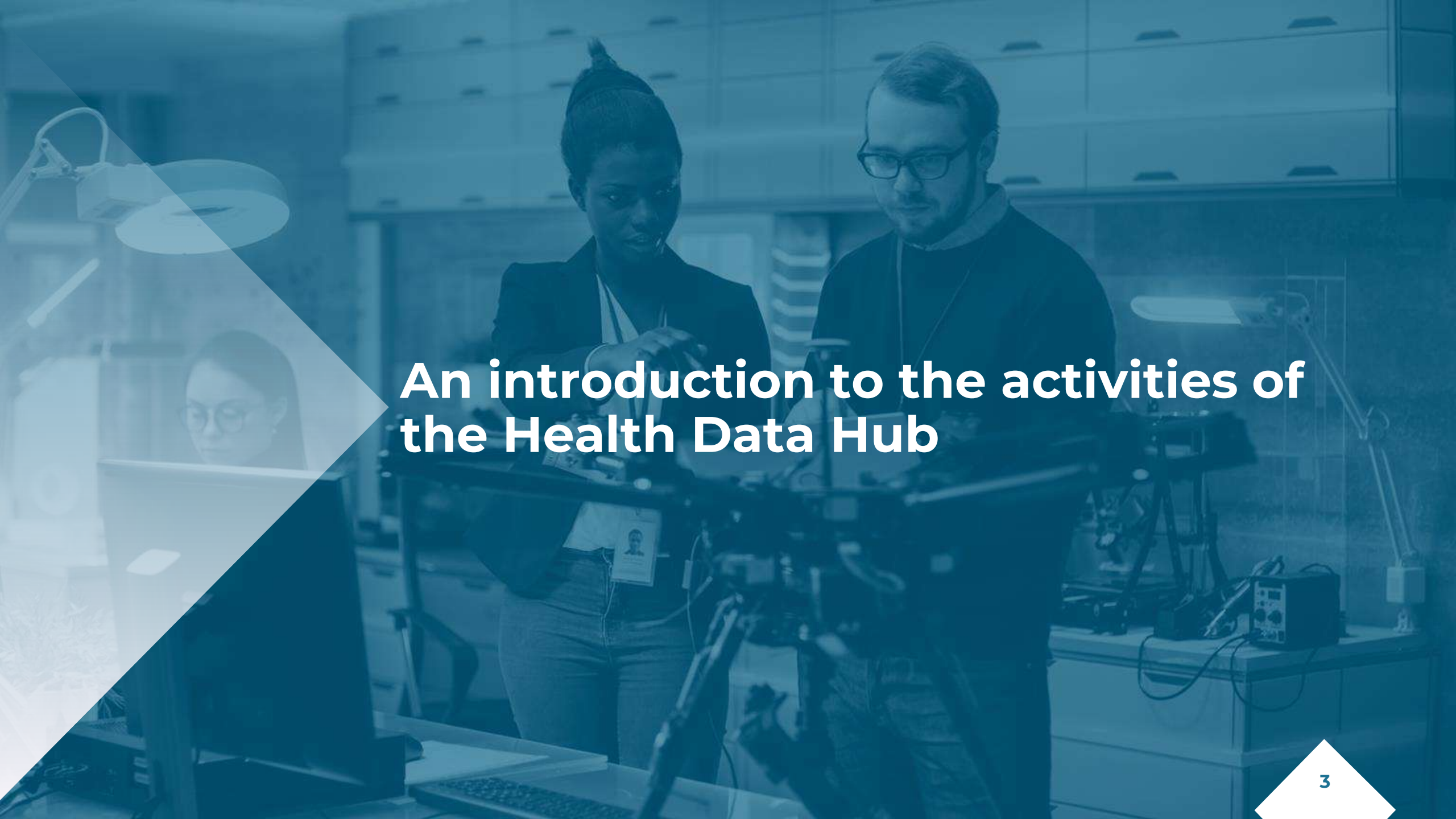
An overview of the National Health Data System and the data warehouse network in France

I. **The Health Data Hub : supporting data integration, analytics and AI (artificial intelligence)**

An overview of the technological platform and practical project examples

I. **How the HDH is helping to shape the landscape of secondary reuse of health data in Europe**

Overview of the HDH's involvement within the EHDS



An introduction to the activities of the Health Data Hub

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 176 innovative projects to this date.

The Health Data Hub is a public body tasked with facilitating access to health data for projects in the public interest

As unique gateway the HDH provides support to prospective data users and transparency to citizen



Requesting access

- The HDH is the one-stop shop for data access.
- **Data access applications** are submitted online to the HDH through a **dedicated platform**.
- The HDH verifies the completeness of the file and transmits it, if necessary, to the Ethic and Scientific Committee for evaluation.
- **The HDH does not play the role of data permit authority** today.
- Local data access is enabled by other actors, eg. through hospital data warehouses incl. local access committees.



Transparency on all projects

- Transparency promise: **All projects are published on the [public project register](#)**. Also for projects not requiring an authorization ("MR")
- A standard information template for each project, including:
 - Context, objectives, study population and methodology
 - Data sources mobilised
 - Involved actors
 - Timeline of the project
- The HDH publishes some **key figures** around data access.

#6000
registered projects



Train and support

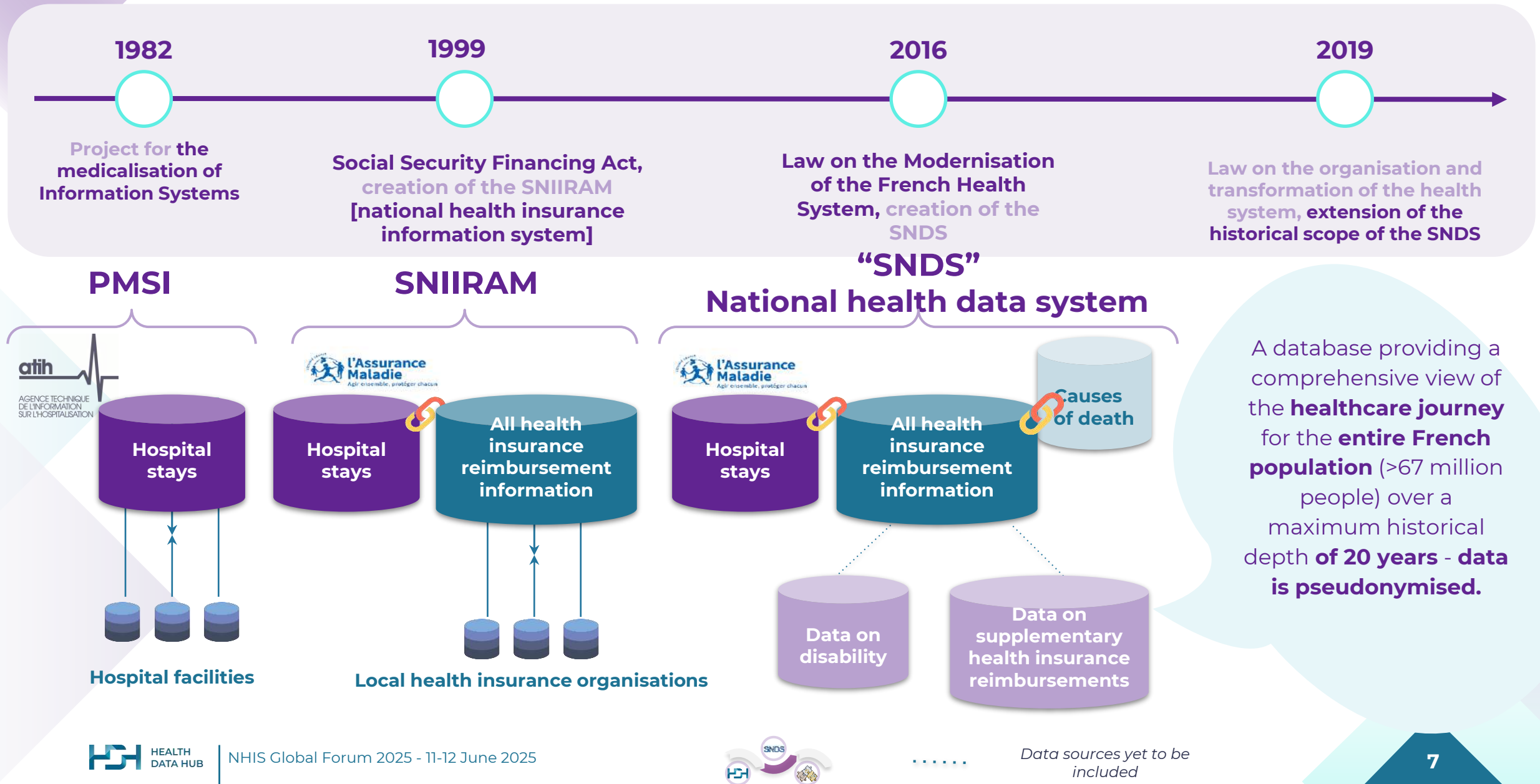
- Regular **trainings** for data users
- **"Starter kit"**, including templates and pedagogical documentation to guide project leaders through the process.
- **Support desk** service to the user community to follow regulatory requirements





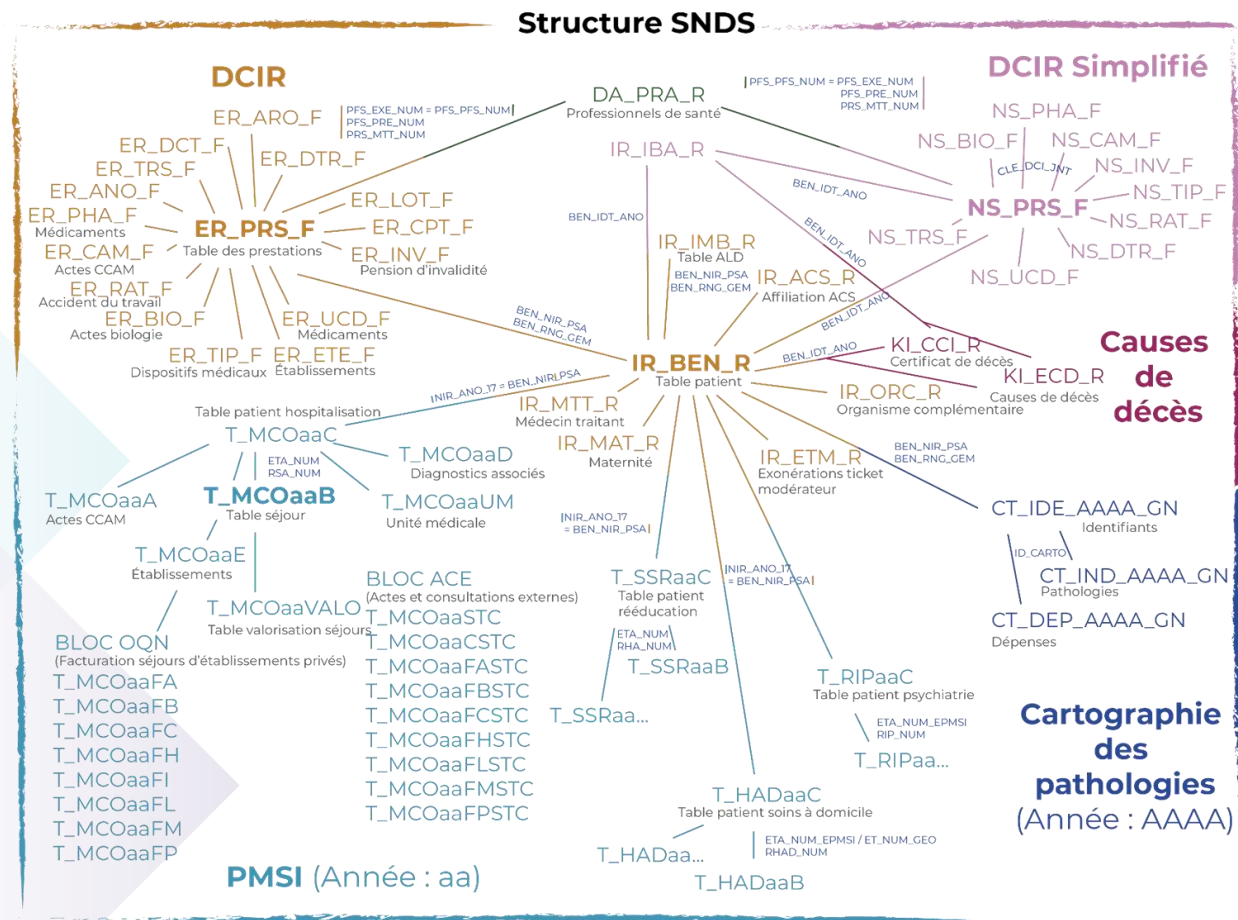
The French digital health infrastructure for secondary data reuse: An overview of the National Health Data System

What is the expanded SNDS ? Context of creation

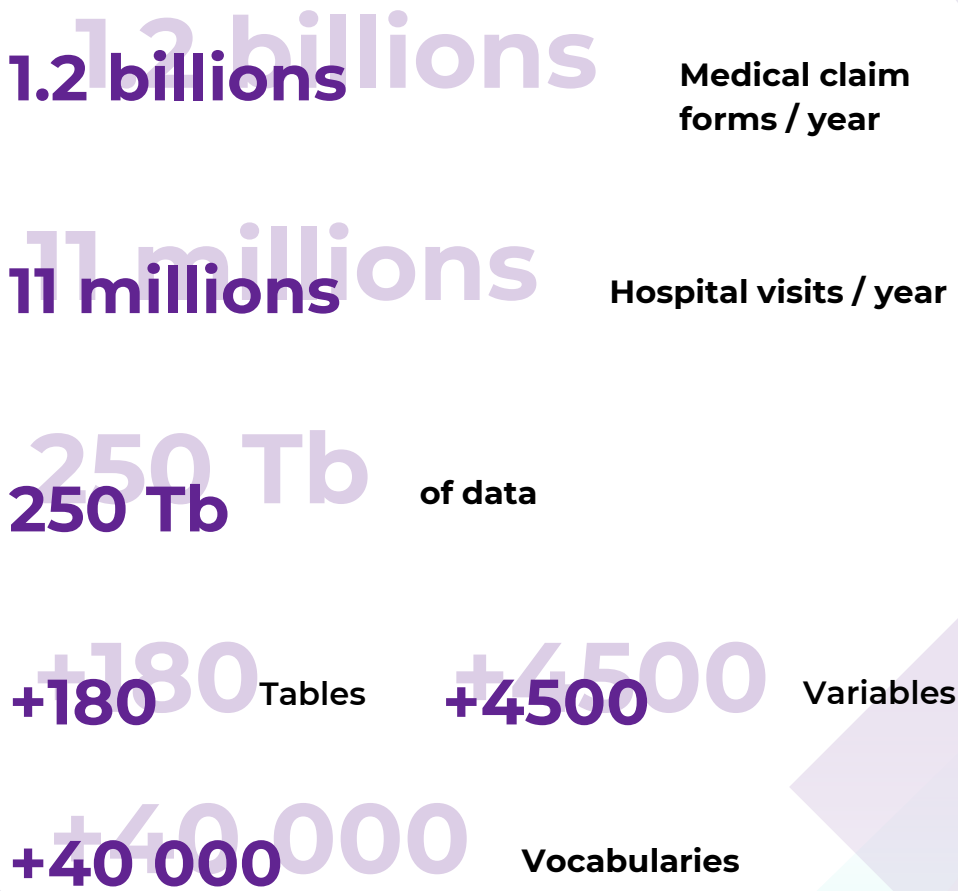


Key figures for the SNDS main database

Structure of the SNDS main database



Key figures



Information available and not available in the main SNDS database

Information on patients, beneficiaries

Age, gender, postcode, type of scheme, long-term illness, disability or occupational disease, date of death, medical cause of death, address at the time of death, doctor.

Healthcare consumption

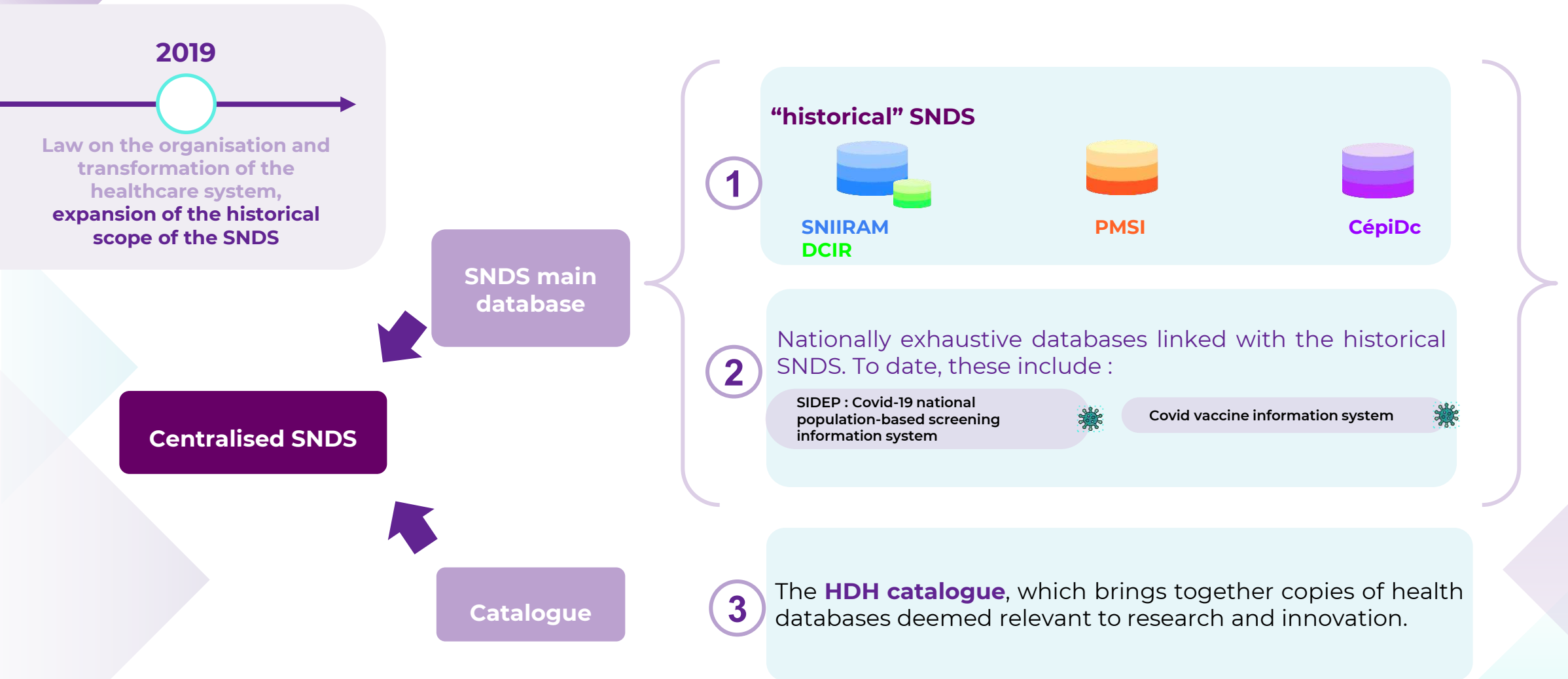
Outpatient care consumption (dates, drugs, visits to healthcare professionals, rehabilitation, nursing, etc.)
Inpatient care consumption (public hospitals, private clinics, some drugs, procedures, medical equipment, medical units, etc.), provider specialty

Long-term illness, disability or occupational disease, Diagnosis, causes of death, drugs, procedure

Information about treated pathologies

-  results of biological tests
-  diagnostics for outpatient medicine
-  risk factors (e.g., smoking, alcohol, etc.)
-  non-reimbursed drugs
-  drugs for in-hospital patients (e.g., antibiotics, pain killers, etc)
-  consistency of the treatment with the patient's condition
-  stages of disease severity
-  care consumption by elderly residents in institutions
-  socio-economic characteristics

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



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.

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 supported in their construction



P4DP: creation of a data warehouse for community care

ISIS, REALIGIST: creation of an oncology reference database

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



Databases from call for project



ARAC: remaining expenses

APSOREN, TARPON: emergency medicine, traumatology

SEDAAR: ophthalmology

APRIORICS, PRECISION PREDICT,

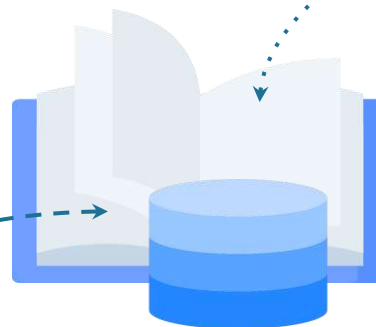
NetSARC: oncology

Other relevant databases



INTEGRA: hospital cohort on metabolic diseases

UroCCR, biological clinical registry for kidney cancer



SNDS
Système national des données de santé

The Hydro project example : predicting Heart Failure crisis to prevent aggravations and reduce hospitalisations



Context

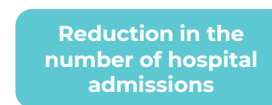
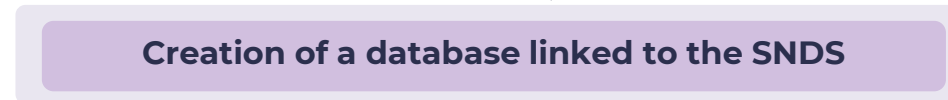
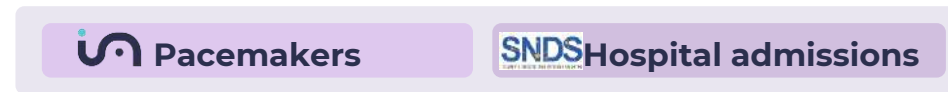
The HYDRO project aims to **predict heart failure crises leading to hospitalisation** for patients with pacemakers in a context where **existing solutions are generally not predictive**

45 000 data from pacemakers

3,8 M French patient data from the SNDS



Approach and first results



Support from the HDH and CNAM in carrying out the linkage (improvement of the rate from 50% to 85%)

Development of the algorithm on the HDH technical platform with the support of the HDH



the medical device created thanks to the HDH has been certified by the Food and Drug Administration (FDA)

The Combo projet example (evaluating new COMBinations in Oncology with real-world data)



Context



In oncology, it's no longer a question of finding a miracle molecule, but of **identifying the best combination of existing molecules** to develop new products.



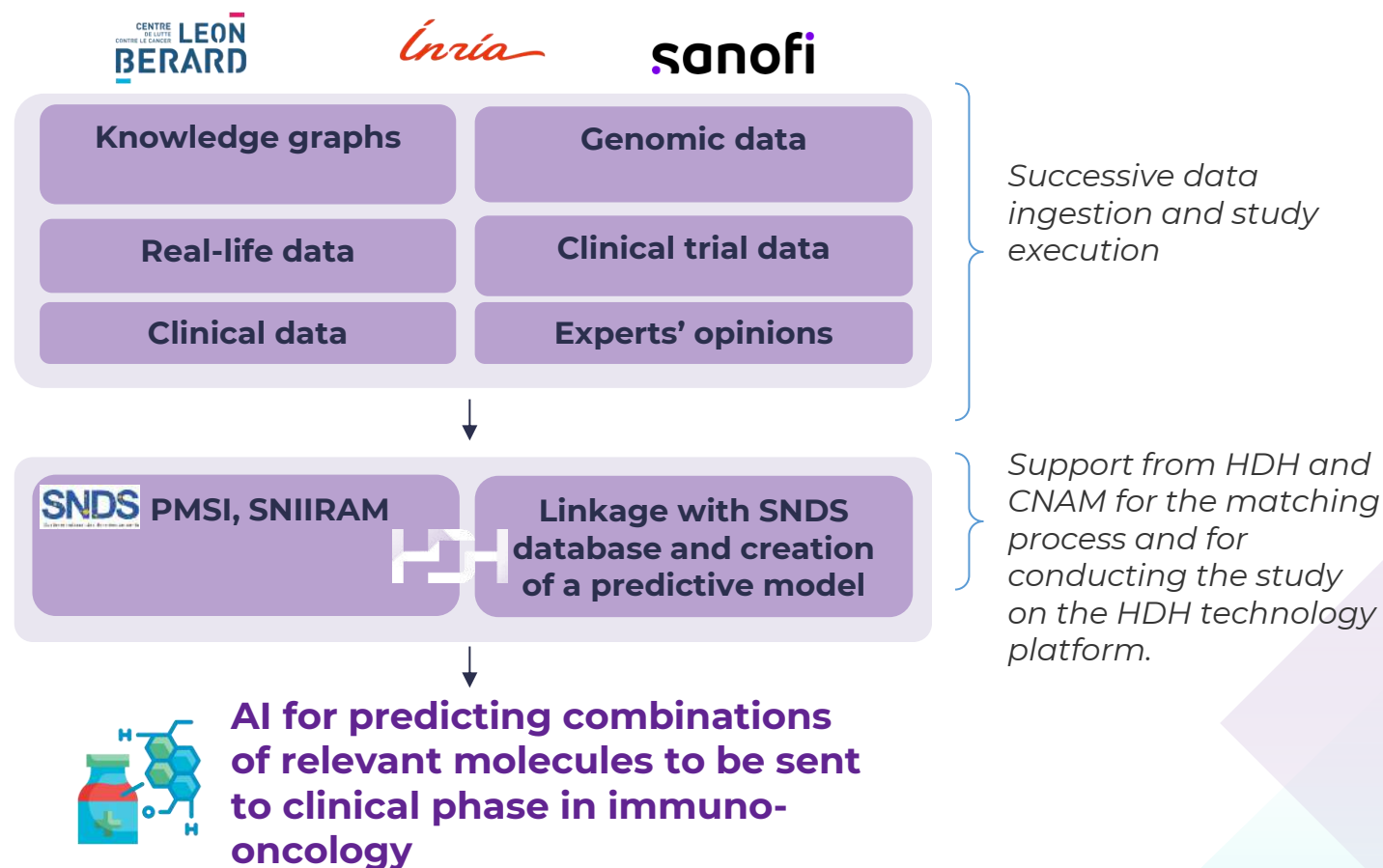
Development programs are therefore **extremely costly**, due to the large number of possible combinations to be tested. **Many of them fail in the pre-clinical and clinical phases.**



The **COMBO project** aims to **develop an algorithm for predicting which drug combinations are most likely to be effective and safe in Phase I clinical trials.**



Approach and first results



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



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**.

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



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



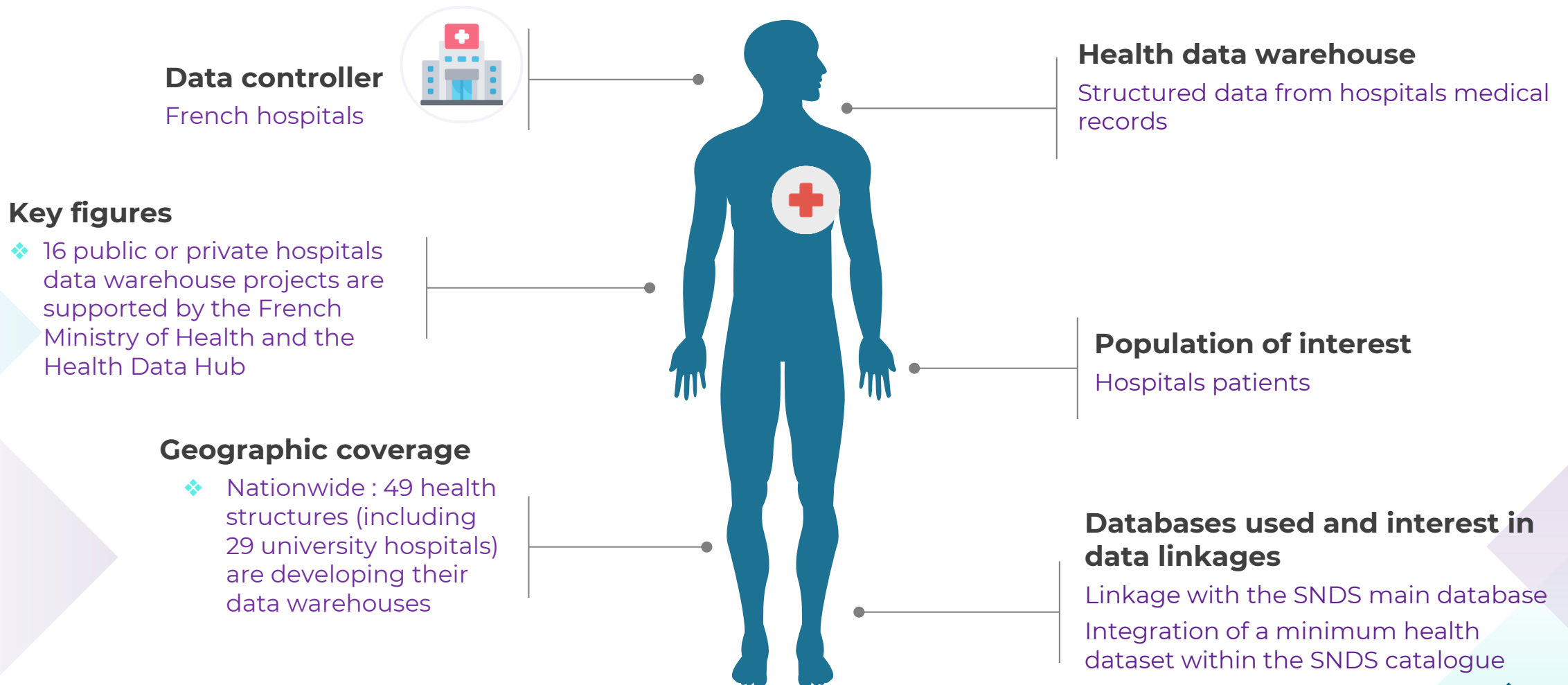
Sustainable financing of health databases and fees for the provision of healthcare data



Standard data provision contract

The example of the hospitals data warehouses

Nationwide network of hospital data



BactHub projet (impact of antibiotic exposure on the risk of community-acquired resistant bloodstream infections)



Antimicrobial resistance (AMR) is a major public health problem: ~ 5M deaths associated with / ~1.1M attributable to AMR worldwide.

Dissemination in the community

Murray et al., Lancet 2022 2024



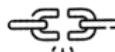
Antibiotic consumption contributes to AMR. But **which antibiotics, and what about community acquired infections?**



The BactHub project aims at **quantifying the impact of antibiotic exposure on the risk of community-acquired resistant bloodstream infections**

L. Watier, S. Abbara

Clinical data-warehouse of the AP-HP (14 hospitals, 2017-2019, Paris area)

Matching:
98%


Main database of the SNDS (health insurance claim data)

**Precise clinical and microbiological data
Complete exposure to antibiotic and healthcare
Precise identification of community-acquired infections**

↓
Administrative, technical and human support of the HDH + analysis on the HDH platform

> 30 000 bloodstream infections among which
10 000 are community-onset and
5 000 are community-acquired (Friedman criteria)

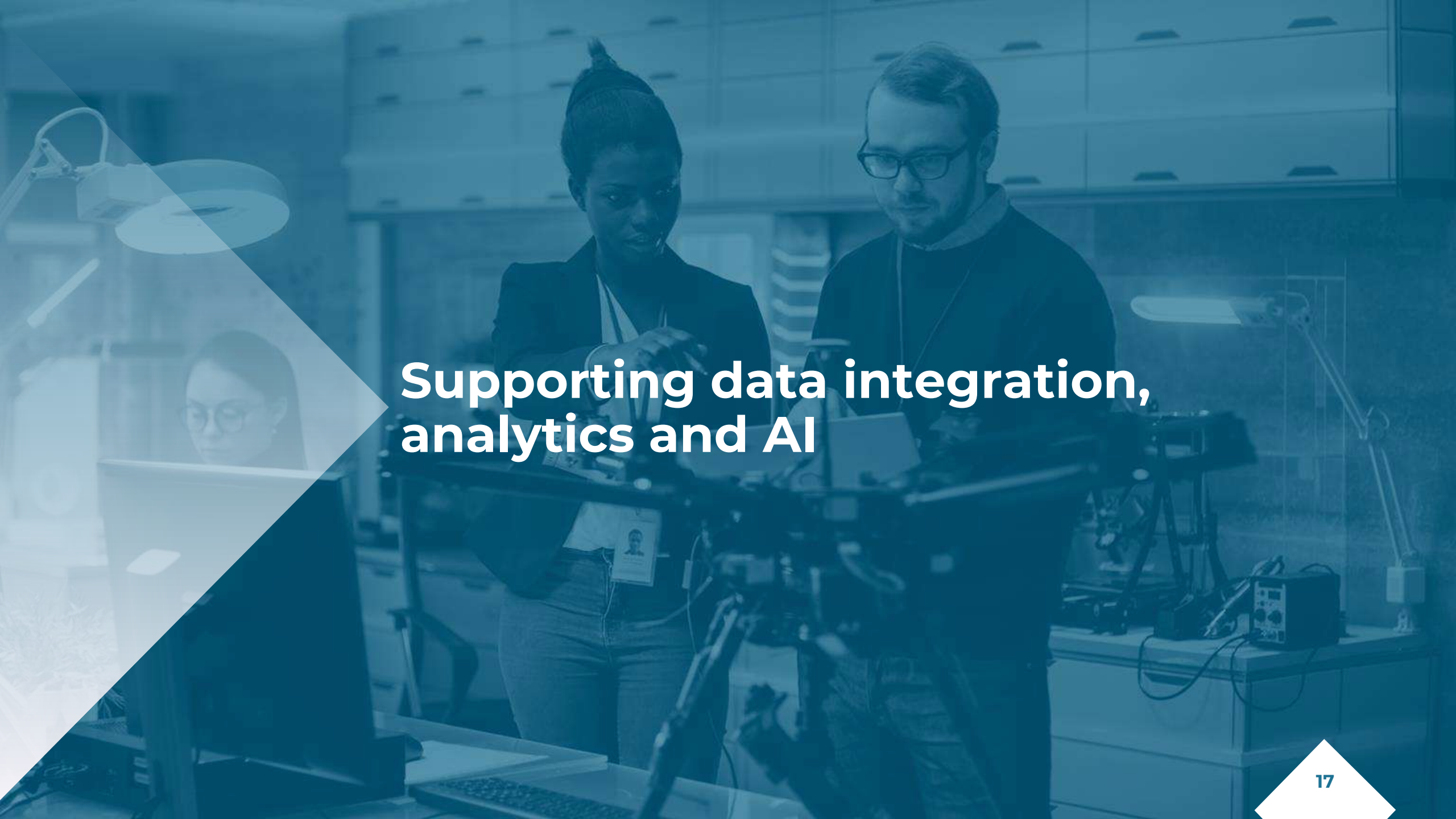


Study 1



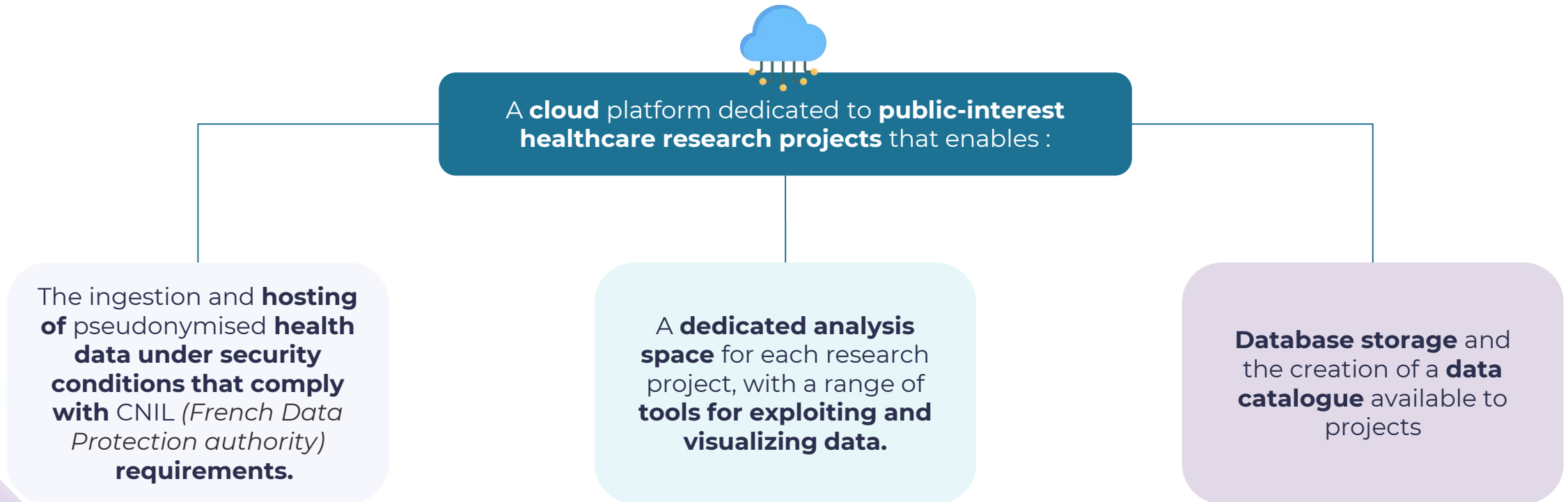
Study 2

Ongoing: link between antibiotic exposure/ 3GC-R *E. coli* community-acquired infections



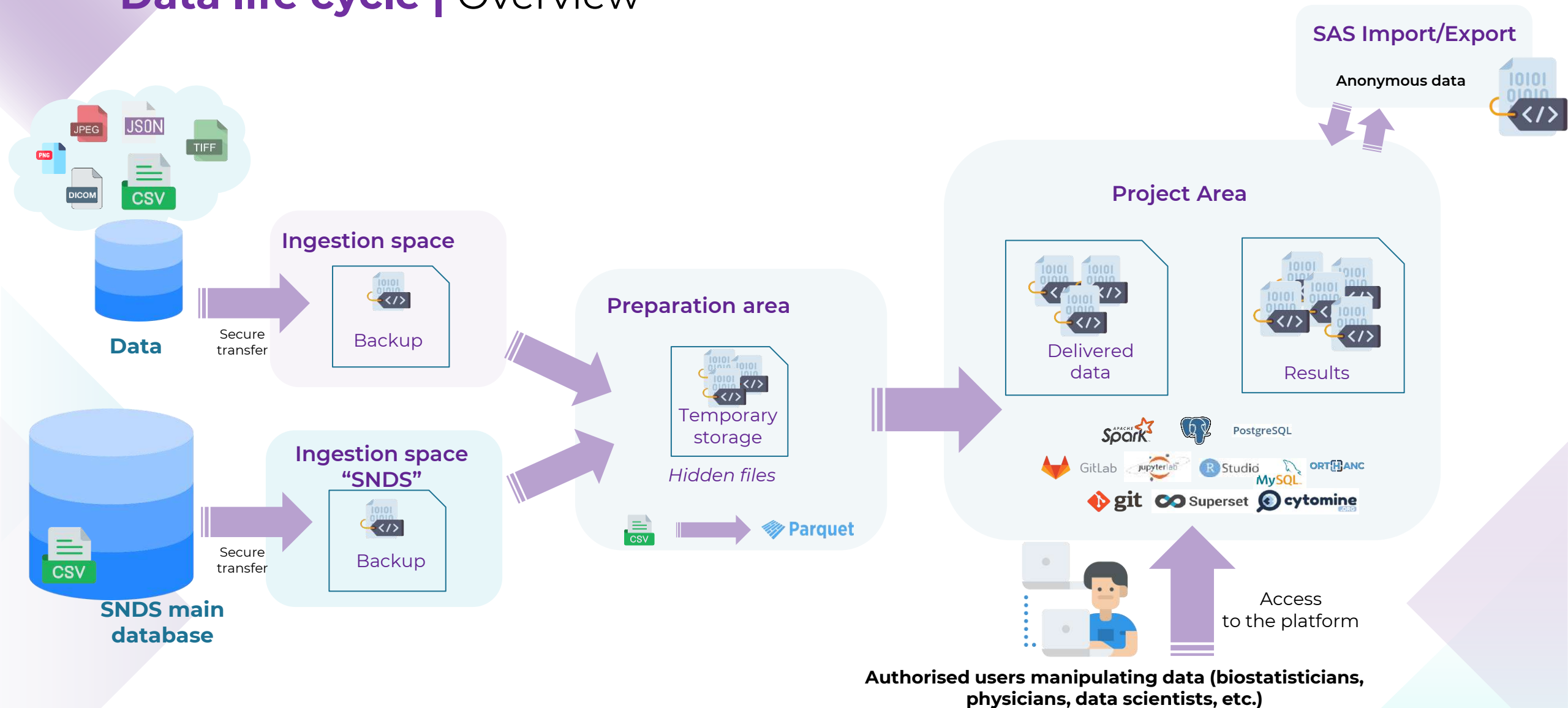
Supporting data integration, analytics and AI

What is the technological platform? | Focus



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.

Data life cycle | Overview



Authorised users manipulating data (biostatisticians, physicians, data scientists, etc.)

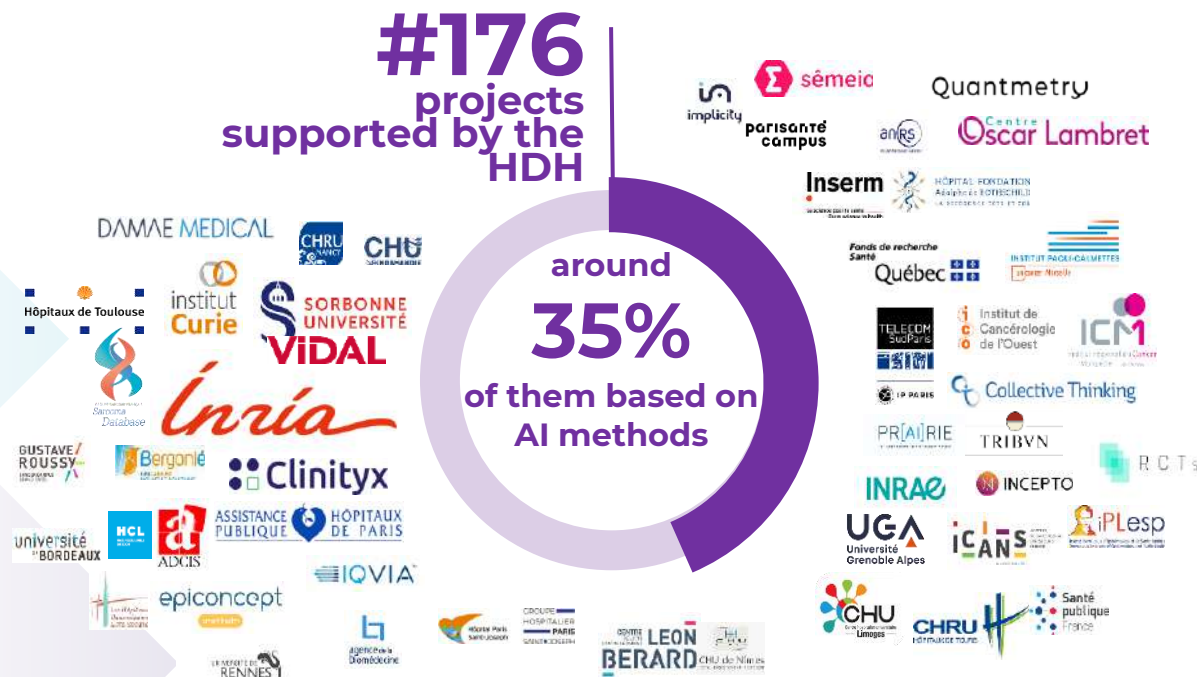
Users can be in different structures if they are partners in the project (also in Europe).

3. Technological offer | current offer*

	Programming languages and IDEs	Big Data infrastructures	Data visualisation	Data storage	Images	Collaboration
Available tools						
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

The HDH makes its services available to support innovative projects using public interest AI approaches



HDH has the capacity to support several projects in public-interest AI

More than a third of the projects supported by the HDH use artificial intelligence tools, and 33 of these are led by research or higher education organizations.

Deep-piste project to evaluate the contribution of AI to organized breast cancer screening

Survival GPU to mobilize graphics cards to accelerate survival analyses on SNDS main database data

CombO to predict the most effective drug combinations in oncology

A few examples of projects using artificial intelligence that delivered initial results

Diagnostic support

NHANCE



Description

Build up a database of non-nominative ultrasound images and make the 1st AI models more reliable.

Results

- 80k semi-automatically pre-processed images
- Publication of two scientific articles ([1](#), [2](#))
- A confirmed trial of automated localisation of focal liver lesions

INNERVE



Description

To build the first AI-based solution dedicated to counting small nerve fibres.

Results

- Tool developed and in deployment phase
- Can be used in a research context
- Retrospective study in progress

Prevention

HYDRO



Description

Using care pathway data to automatically annotate telemonitoring data

Results

- Algorithm under development
- Application submitted to the FDA for marketing

TARPON



Description

Cross-reference patient information on traumatic events with pre-trauma drug consumption data

Results

- Algorithm developed and implemented in a powerful server dedicated to AI
- 97% of CRs classified correctly
- A scientific article ([1](#))

DEEPSARC



Description

Defining the best sarcoma patient pathway through retrospective data analysis (2010 to 2017)

Résultats

- Data linked and transferred
- Initial results produced by statistical analysis
- Description of current analysis population

Data Challenge

TissueNet



Description

Improving medical diagnosis using artificial intelligence to detect precancerous and cancerous lesions of the cervix.

Results

- 574 participants from all over the world
- 20 data-providing centres
- 90%+ accuracy in detecting cervical lesions

VisioMel



Description

Targeting patients with a high risk of recurrence so that preventive treatments can be introduced to improve patient survival

Résultats

- Three winners rewarded - their algorithms will be shared as open source
- Results of the Data Challenge still being analysed

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

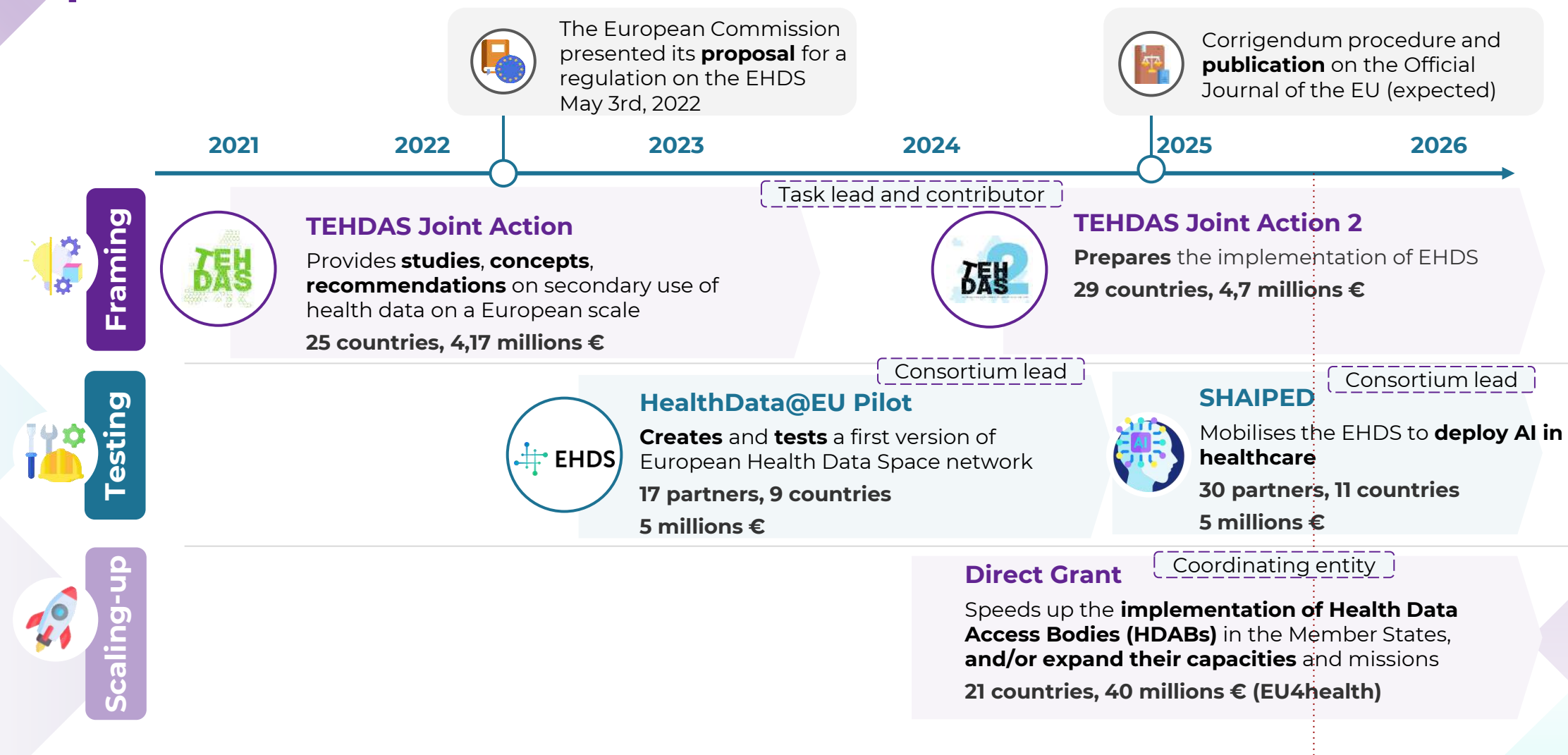


Methodological guides for algorithm refinement (*fine-tuning*) by hospital staff using their own data



How HDH is helping to shape the landscape of secondary reuse of health data in Europe

Growing initiatives for European Health Data Space implementation



Overview of the SHAIPEd project

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

4M€
in EU funding

3 years duration

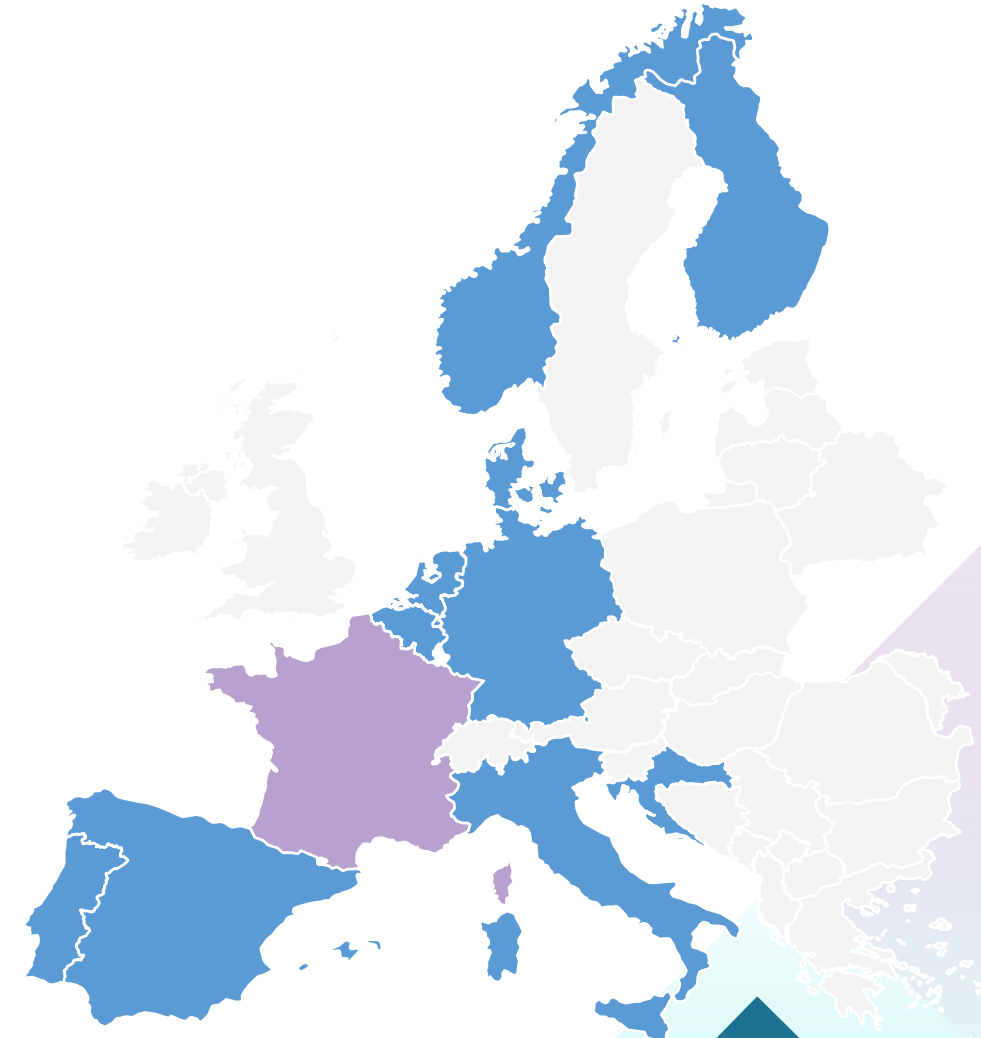
30 Partners throughout Europe

11 Member States involved

Research Institutions / Hospitals / Public Instances / Start-ups

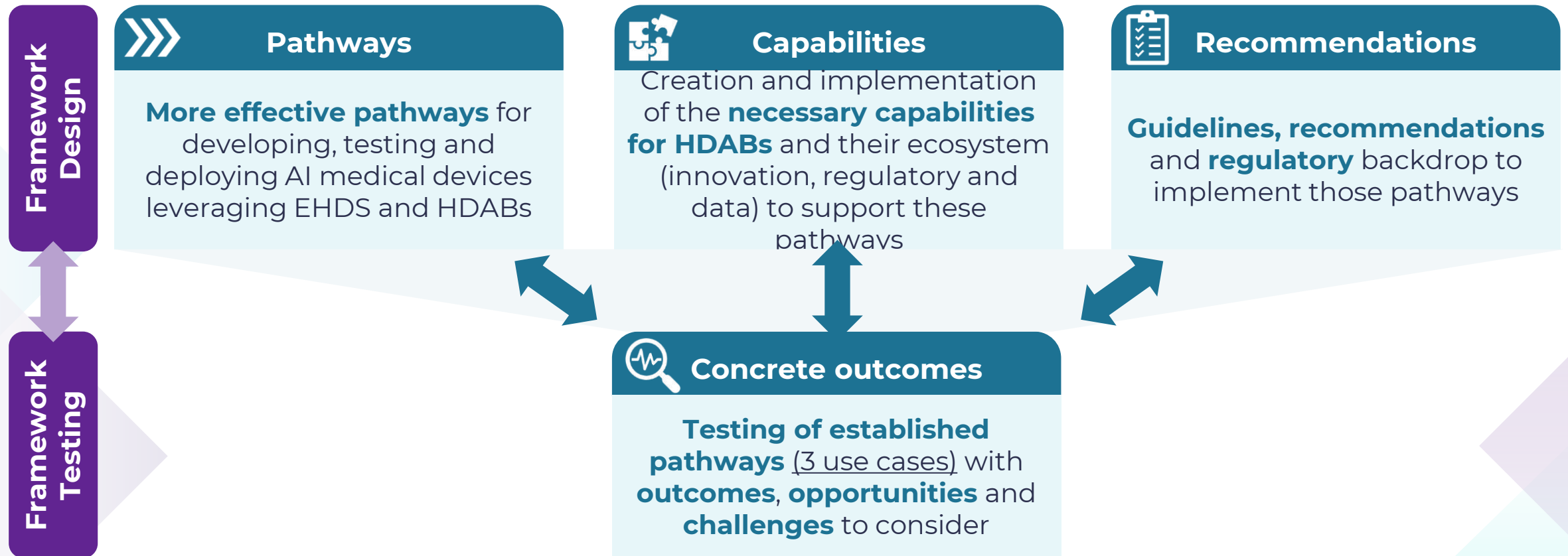


11 HDABs and other related organizations



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





Conclusion

After five years, where do we stand?



Production launch of a technology platform in just one year



Rapid recruitment of staff to reach 116 people by 1 January 2025



Support for 170 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**



176
projects
supported by
the HDH,
of which 123
use the
technology
platform

35

projects in the regulatory phase
(data access request)

33

projects in the data preparation
phase

57

projects running on the
platform by the end of 2025

23

projects completed, or with
intermediary results

24

Calls for project



Do you have any questions ?



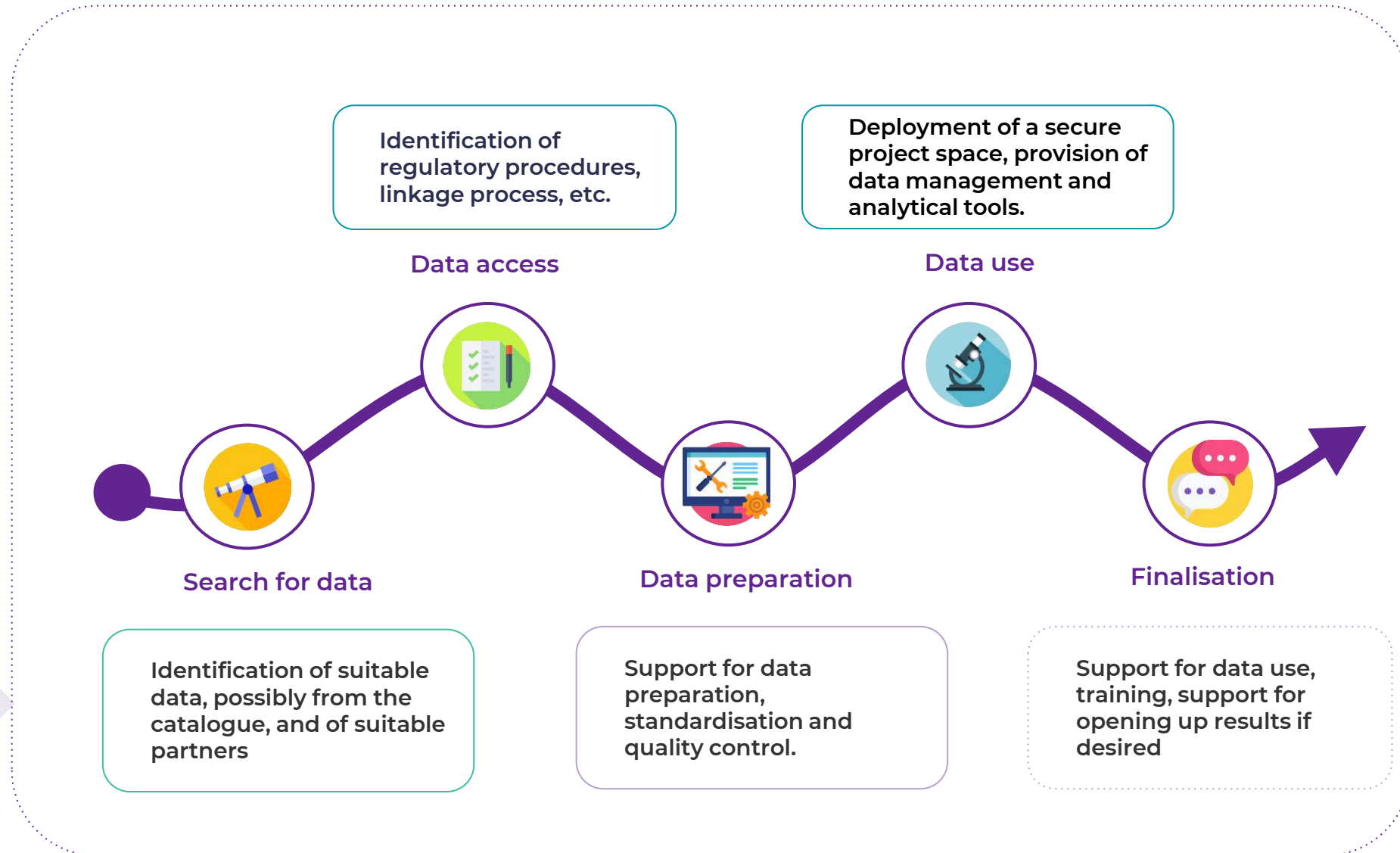
Follow us on social media!



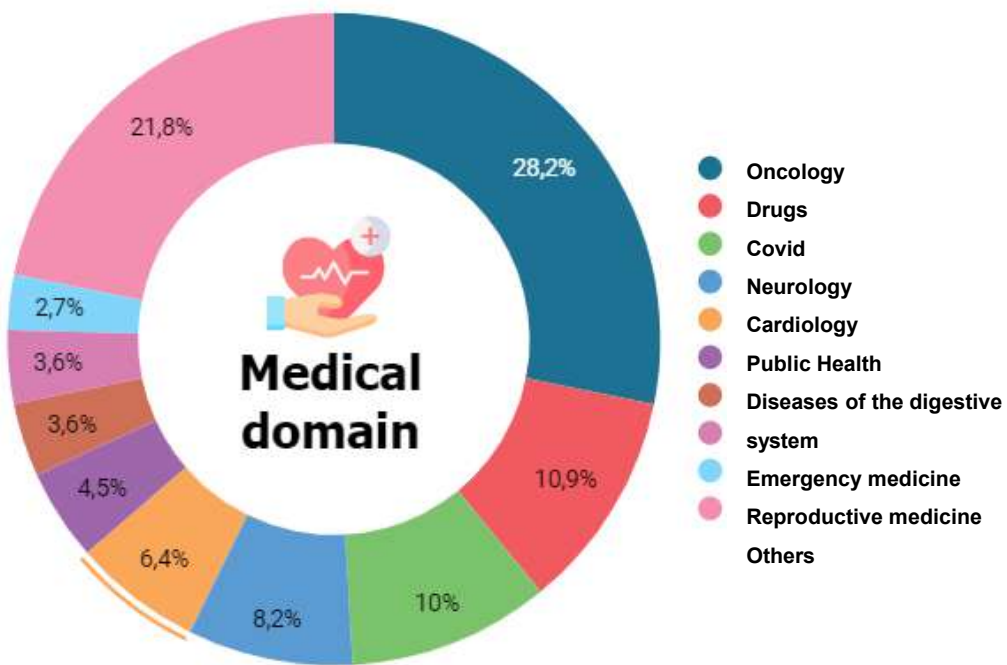


Annex

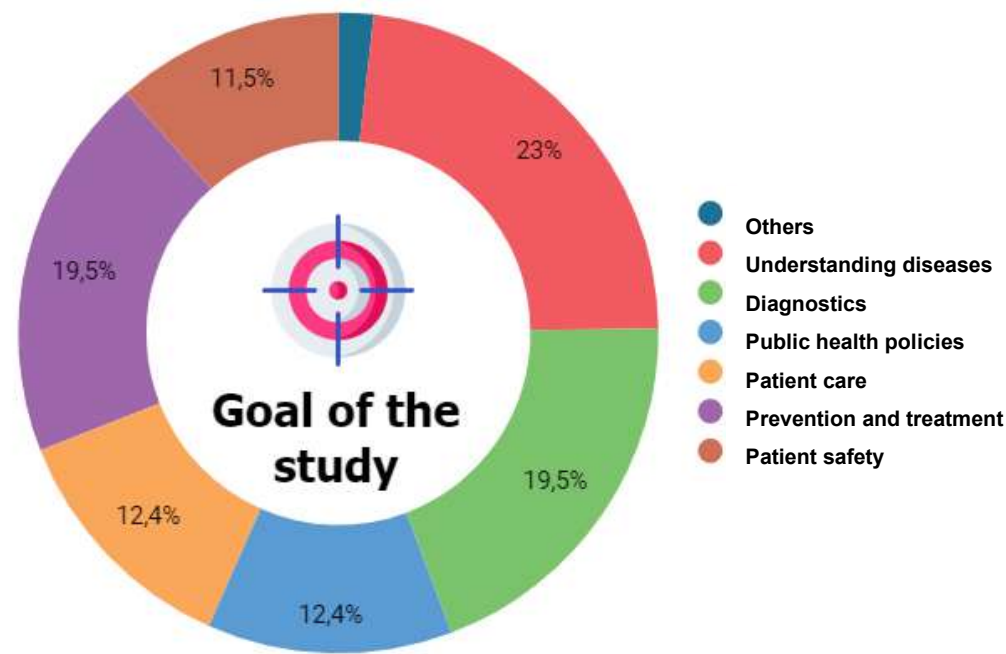
In a nutshell | a comprehensive service offer to guide projects



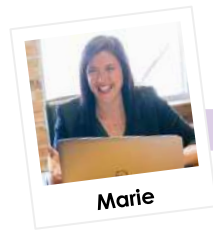
In a nutshell | projects addressing a multitude of areas and objectives



HDH responds to a **real need in the ecosystem**, which is the subject of significant demand that will accelerate with the promotion of the use of healthcare data for a variety of purposes.



Requesting access to data | The support of the HDH



Marie



Health data



without
starter kit

with
starter kit



The HDH is the **one-stop shop** for data access. Data access requests are made online via a dedicated platform. It provides a **starter kit** to guide applicants throughout the process.

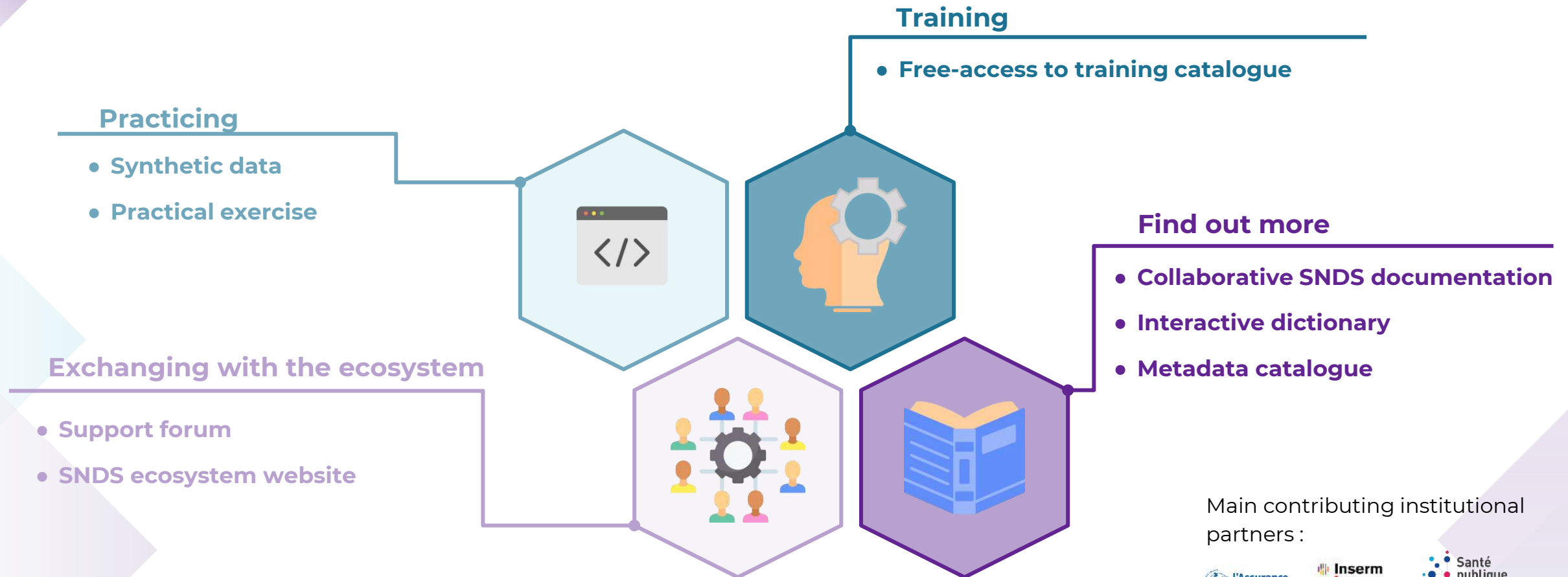


The HDH checks the completeness of the application and, if necessary, forwards it to the **Ethics and Scientific Committee (CESREES) for evaluation**. The committee provides an opinion within one month on the research's **objective** and **methodology**; the need for the use of personal health data; **ethical** relevance; **scientific quality** of the project; and the **public interest** of the project.



If CESREES gives a favourable opinion, and at the applicant's request, the **CNIL** (data protection agency) evaluates the data access request for approval within a two-month period, which can be extended once.

The Health Data Hub makes available open resources to facilitate using and understanding the SNDS main database



Main contributing institutional partners :



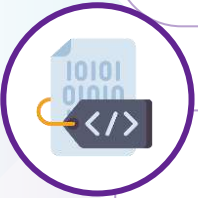
Equivalent resources are available on the CNAM portal once access has been granted.

Increasing findability of databases added to the SNDS catalogue : the data directory



The HDH lists on a page the databases that are added to the SNDS catalogue along with their descriptions.

This information is available on the [HDH website](#).



Beyond the directory, the HDH develops a **national metadata catalogue** to make it easier for project leaders to identify the data available, its content and its quality.

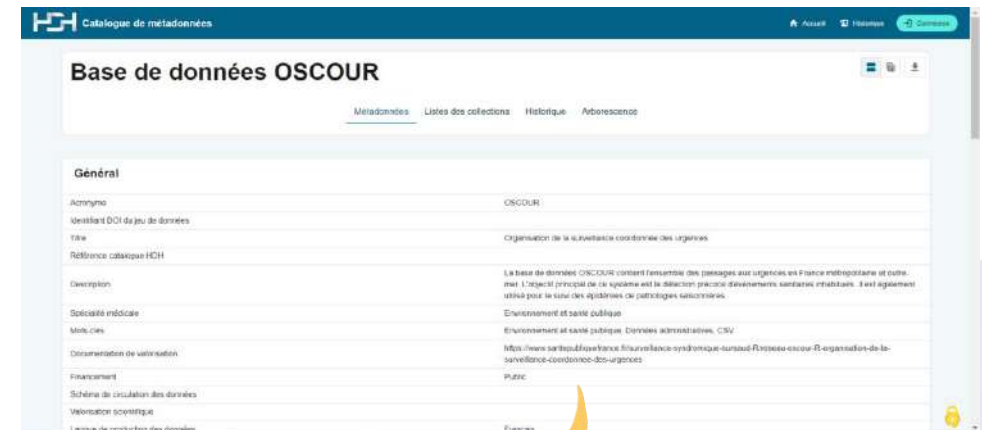
The screenshot displays the 'Catalogue des données' (Data Catalogue) on the Health Data Hub website. The header includes navigation links for 'Vous êtes...' (Citoyen, Utilisateur de données, Responsable de données, Candidat, Start-up) and 'EXPLORE LE HDH', 'DÉPOSER VOTRE PROJET', 'NOTRE OFFRE DE SERVICE', 'DÉCOUVRIR LE CATALOGUE', and 'ACTUALITÉS'. The main content area shows a search bar with the text 'Rechercher un intitulé' and a search icon. Below the search bar, there are filters for 'Domaine Médical (Catalogue des données)' and 'Catégorie de données'. The search results list several databases:

- e-MUST** - Évaluation en Médecine d'Urgence des Stratégies Thérapeutiques des infarctus du myocarde de moins de 24h pris en charge par les SAMU/SMUR d'Île-de-France. Domaines: MÉDECINE D'URGENCE. Catégories: INFORMATIONS RECUEILLIES À L'OCCASION D'ACTIVITÉS DE PRÉVENTION, DE DIAGNOSTIC, DE SOINS OU DE SUIVI SOCIAL ET MEDICO-SOCIAL.
- e-SIS** - Base de données du dépistage organisé du cancer du sein (DOCS) du Gard et de la Lozère entre 2004 et 2019. Domaines: CANCÉROLOGIE. Catégories: INFORMATIONS RELATIVES AUX CONDITIONS SOCIALES, ENVIRONNEMENTALES, AUX HABITUDES DE VIE ET AU CONTEXTE SOCIO-ÉCONOMIQUE DES PERSONNES CONCERNÉES.
- OSCOUR** - Organisation de la surveillance coordonnée des urgences. Domaines: MÉDECINE D'URGENCE. Catégories: INFORMATIONS RELATIVES À LA PRISE EN CHARGE SANITAIRE, MEDICO-SOCIALE ET FINANCIERE ASSOCIÉES À CHAQUE BÉNÉFICIAIRE.

Giving more visibility to strategic datasets with a structured metadata catalogue

A granular metadata catalogue, tailored to healthcare data and part of a European approach

- **Metadata structured in three levels** (dataset, collection, database (structured database, set of images))
- **Use of standard ontologies** (e.g. MeSH: Medical Subject Headings)
- **Alignment with European programmes** (TEHDAS recommendations, creation of a DCAT-AP extension for the medical field)
- **Designed for the healthcare ecosystem** (used to document SNDS catalogue databases, white-labelled service provided to future healthcare data warehouses, source code ultimately open sourced)



MVP in production and available at this address:
<https://catalogue-metadonnees.health-data-hub.fr/>

Permissible and prohibited purposes

All studies, research, and evaluations involving personal health data must pursue a public interest purpose to be authorised (Art. 66 loi Informatique et libertés).

The main SNDS database also has its own specific purposes :



Potential authorised purposes :



- health information (healthcare provision, medical and social care and quality) ;
- definition, implementation, and evaluation of health and social protection policies ;
- understanding of healthcare expenditure, health insurance and medico-social expenditure ;
- information for professionals, organisations, and healthcare/medical-social establishments about their activities ;
- health surveillance, monitoring, and safety ;
- research, studies, evaluation, and innovation in the health and medical-social fields :

Prohibited purposes :



- promotion of health product to healthcare professionals or establishments ;
- Or the exclusion of cover from insurance contracts or the modification of insurance contributions or premiums for an individual or a group of individuals.

To demonstrate that a project is in the public interest, it is possible to rely on the project's aims, benefits, transparency and scientific integrity.

Public interest is not inherently incompatible with commercial interest, and the quality of the applicant is not a criterion.



Follow us on social media!

