ANNUAL REPORT 2021

HEALTH
DATA HUB

CONTENTS

1

2

3

Bruno Maquart,
President of the
Health Data Hub

PREFACE

The Health Data Hub's team of nearly sixty employees, with a target of reaching one hundred staff set for the end of 2022, is now fully equipped to fulfil the missions entrusted to it by law and carry out its projects, by mobilising multiple players. The Hub can be proud of its first successes.

The projects supported by the Hub are starting to produce results. Examples include the BACTHUB project by INSERM and AP-HP, which aims to create a unique database on blood-borne bacterial infections, the NHANCE project by INRIA and AP-HP, which is based on a database of non-identifying ultrasound images of the abdomen, and the INNERVE project by Quantmetry, which aims to develop software for the early diagnosis of small-fibre neuropathies. Some of them have already been published.

In just two years, the Health Data Hub has carved out its place in the health data ecosystem, serving both collectors and users of health data. A few steps still need to be taken, notably making the primary database of the main French health insurance system (L'Assurance Maladie) accessible to all, bearing in mind that the Hub has been jointly responsible for its processing since the publication of a decree last June. The Hub has been working hand in hand with the French National Health Insurance Fund (Caisse nationale d'assurance maladie – CNAM) for several months, which has improved its teams' expertise. I would like to take this opportunity to thank the CNAM, which, alongside the Hub, is more committed than ever to meeting the community's expectations.

In just two years, the Health Data Hub has carved out its own place in the health data ecosystem

The Health Data Hub has recently gained international recognition. Europe is eager to promote data flows between its Member States, and is stepping up its initiatives. The Health Data Hub, together with its national partners, is contributing actively to this process. For example, it has brought together a consortium of some fifteen major European players, including the national health data platforms of several Member States, capable of responding to the European Commission's call for applications in preparation for the creation of a European Health Data Space.

Finally, the Health Data Hub is a resolutely citizen-oriented project. Health data arouses considerable debate, which makes it essential for citizens to have their say. At the same time, it is important to reiterate the need to promote the use of a shared asset, while upholding individuals' rights, in order to take up the challenge of providing the most effective medical care for all, while fully exploiting the power of new technologies.

I hope you enjoy reading this report.

EDITORIAL

Stéphanie Combes,
Director
of the Health Data Hub

How can we improve screening and ensure that patients are treated as promptly as possible? How can we offer them the best treatment throughout their lives? How can we support health professionals in an increasingly complex clinical context or in response to a health crisis? **Thanks to artificial intelligence, health data is now opening up new opportunities and enabling major advances.**

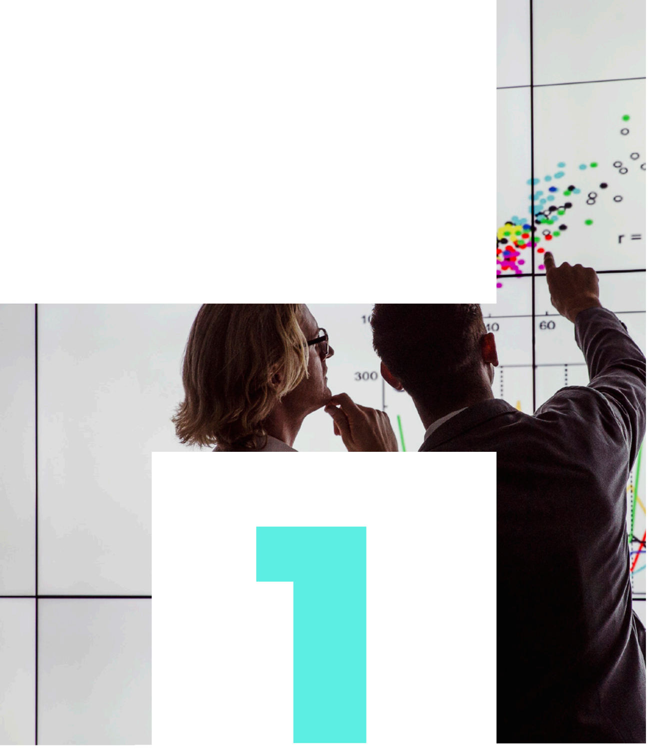
It has even become indispensable. Data from cardiac implants can prevent heart failure; data from mammograms can speed up breast cancer screening; data on care pathways can improve the management of rare diseases, etc. In France, we are fortunate to have highly detailed databases, putting all these innovations within our reach.

Meeting this challenge was the reason for creating the Health Data Hub two years ago: to **provide simple, unique, transparent and secure access to a catalogue of health databases**. Our technological platform enables the leaders of public-interest projects to make use of extensive data sources, cross-reference them and process them, using massive computing power to run complex search algorithms. The Health Data Hub is on hand to support their projects and help them find tomorrow's keys to improving the health of all citizens.

2021 – a year of consolidation for the Health Data Hub

By the end of 2021, the Health Data Hub and its 56 employees were already supporting 55 projects. In addition to these examples of data use, the HDH has increased its number of partnerships with key players in the ecosystem in order to scale up its systems and support innovation. New calls for projects have been issued, strategic activities have been launched and a scientific roadmap is taking shape. Following the decree formalising the extension of the French National Health Data System (Système National des Données de Santé – SNDS) on 30 June, the Health Data Hub and its partners have been working tirelessly to satisfy all the requirements for implementing the sharing of the core database and an initial version of the data catalogue via the platform.

Resolutely open to civil society and the international community, the Health Data Hub is fully committed, together with its European counterparts and citizens, to the projects that are laying the foundations for the future European Health Data Space. The inauguration, last December, of PariSanté Campus, of which the Health Data Hub is a founding member, bears witness to our commitment and ambition, in the interest of our country's health sovereignty.



THE HEALTH DATA HUB – KEY INFRASTRUCTURE FOR HEALTH RESEARCH

FOUR PUBLIC-INTEREST MISSIONS

Established by the Law of 24 July 2019 on the organisation and transformation of the French healthcare system, the HDH is a public entity that gives stakeholders easy access to non-personal health data, hosted on a secure platform that respects citizens' rights. Project leaders can cross-reference data and analyse it to improve the quality of care and patient support.
To do this, the HDH carries out four missions.

1. **Provide a single gateway** to assist project leaders with their administrative procedures. These projects are examined by the French Scientific and Ethical Committee for Research, Studies and Evaluations in the Health Sector (CESREES), which issues its opinion before they are submitted to the French Data Protection Authority (CNIL) for authorisation;

2. **Open up access to a catalogue of partner databases.** Developed in a progressive and iterative manner, the content of this catalogue will be laid down in an order issued after the CNIL has published its opinion, which was under consideration when this report went to press; The databases in the catalogue will be replicated on the technological platform in order to cross-reference them with other databases, in agreement with the actors from whom the data originated;

3. **Provide a secure, state-of-the-art platform** with advanced data storage, computing, comparison and analysis capabilities;

4. **Co-ordinate the ecosystem to accelerate innovation** by promoting the sharing of experience and knowledge.

The HDH's service offering is complementary to that of public or private actors managing administrative databases, registers, cohorts, or surveys. It meets a need expressed by two categories of projects:

- projects aiming to cross-reference different data sources, where such cross-referencing, linkage or matching is complicated to achieve;
- projects requiring large volumes of data, so that the statistical methods used can be considered reliable.

In both cases, the HDH technological platform acts as a trusted third party by securely bringing together the different data sources and enabling their exploitation using state-of-the-art technologies.

Serving the public interest

New tools such as artificial intelligence pave the way towards significant advances in patient care, the assessment and choice of treatments, and the management of health systems. This is achieved by processing and cross-referencing large volumes of high-quality data. Here are some examples.

Accelerating and improving care of ophthalmology patients through the SEDAAR project, led by Dr Rabih Hage MD, Ophthalmologist, Fondation A. de Rothschild Hospital

An estimated 1.7 million people in France suffer from visual impairments. These disorders are mainly related to retinal pathologies. The vast majority of diagnoses are made by ophthalmologists based on a variety of imaging-based tests. Automating some of these diagnoses is an important challenge because the earlier the diagnosis, the greater the likelihood of the effective treatment of pathologies. In addition, specialists are becoming scarcer.

Preventing serious adverse events during the care pathway through the HUGOSHARE project led by Pr Marc Cuggia, GCS Hôpitaux Universitaires Grand Ouest (HUGO)

Every year in France, the misuse of medicines is responsible for 10,000 deaths, mainly among people over 65 years of age. In the future, by analysing patients' complete drug therapy trajectories (in the hospital and non-hospital sectors), it will be possible to prevent accidents related to treatments or to the interruption of treatments, to improve the most fragile patients' care pathways, and to develop an additional component in prescription-support tools for health professionals in order to promote the proper use of drugs.

Developing a better understanding of the determinants of patients' out-of-pocket expenses through the ARAC project supported by Malakoff Humanis

In France, knowledge of patients' final out-of-pocket expenses, after reimbursement by the main French National Health Insurance Fund and complementary voluntary health insurance, is still limited. Indeed, there is relatively little data on the differences in these out-of-pocket expenses between individuals (according to the context of their healthcare consumption, their complementary voluntary health insurance cover, their professional or geographical situation, etc.). The results obtained by cross-referencing data from the main health insurance system and for mutual insurance policyholders will enable the identification of difficulties in accessing healthcare, and help to combat the phenomenon of foregoing medical care.

Automating health professionals' activities in order to free up medical time, through the TAMIS project led by Prof. Yaneck Gottesman, Telecom Sudparis

In 10% to 15% of cases, blood samples require additional manual analysis by a haematologist for the precise characterisation of certain blood cells. Images of blood smears can be exploited in order to train algorithms that can more precisely detect and characterise the different cells of interest. This frees up medical time for haematologists, to the benefit of patients, who will obtain their results more quickly.

Assisting health professionals with the diagnosis of pathologies, through the DeepMap project by Damae Medical

Every year, 10 million basal cell carcinomas are diagnosed worldwide. Their diagnosis is important to avoid unnecessary or belated interventions, and can be carried out by biopsy or by imaging techniques. Images collected by artificial intelligence will be exploited and used to develop tools to assist dermatologists with their diagnoses, provide guidance for less invasive surgical and traumatic interventions, and help monitor the patients' treatment.

ORGANISATIONAL STRUCTURE OF THE HEALTH DATA HUB

The Health Data Hub team

The HDH team is composed of some sixty people who share a common desire to help research and innovation players to accomplish their projects.

Governance

The HDH is a Public-Interest Group whose General Assembly has 56 members divided into nine boards.

- State
- Health insurance fund (Caisse d’assurance maladie)
- Complementary voluntary health insurance organisations
- Research organisations
- Healthcare institutions
- Healthcare professionals
- Agencies
- Operators and the independent public authority
- Representatives of health system users
- Manufacturers in the health sector

One representative of each board sits on the Board of Directors, except for the French State which has two representatives: the Directorate for Research, Studies, Evaluation and Statistics (DREES) for the Ministry for Solidarity and Health, and the Directorate-General for Research and Innovation (DGRI) for the Ministry of Higher Education, Research and Innovation. The CNAM (French National Health Insurance Fund), UNOCAM (National Union of Complementary Voluntary Health Insurance Bodies), INSERM (French National Institute for Health and Medical Research), Limoges University Hospital, UNPS (National Union of Healthcare Professionals), France Asso Santé, Santé Publique France and SNITEM (National Union for the Medical Industry and Technologies) are members of the Board.

The General Assembly and the Board of Directors are chaired by Bruno Maquart, Chairman and Chief Executive of Universcience. Gérard Raymond, President of France Asso Santé, is Vice-President of HDH, reflecting the importance of patient organisations in the governance system.

In January 2021, the HDH established a Scientific Board, composed of 28 experts with complementary expertise: doctors, pharmacists, artificial intelligence specialists, lawyers and ethics specialists, geneticists, epidemiologists, health economists, etc. This board is chaired by Professor Alain Livartowski, an oncologist specialising in thoracic tumours and Deputy Head of Data at the Curie Institute, who provides advice to inform management decisions in a highly competitive international environment. He collaborates closely with the HDH's scientific and medical management teams. On 26 November, seven international members joined the board to help the HDH identify partnerships outside France, and new tools and services. They contribute to the wider dissemination of knowledge within the global scientific community and to the HDH's integration into the European health data landscape.

Primarily, the HDH is publicly funded, and received €20 million per year, on average, during the 2019-2022 period. A large proportion of this funding comes from the Public Action Transformation Fund (FTAP) and the National Health Insurance Expenditure Target scheme (ONDAM).

On 31 December 2021, the HDH had 56 employees working in specialised teams providing end-to-end support for projects (Data Access Team, Partnerships Department, data, platform and product), in strategic departments (Scientific, Medical and Citizen Relations), and in support functions.

FIVE PRIORITIES FOR 2021

In 2021, the Health Data Hub continued to focus on extending its relationships with associations and citizens, and developing partnerships with various public and private-sector organisations, while extending its European and international footprint.

These actions were divided into five priorities:

End-to-end project support. The Health Data Hub's teams, in close collaboration with existing partners such as the CNAM (French National Health Insurance Fund), and prospectively with the CNAV (French National Old-Age Insurance Fund), were fully committed to ensuring the success of each project, from the preparation of applications, their targeting and the management of matching, through to the provision of data. A starter kit was disseminated.

Placing the general public at the heart of our actions. With its focus on building a relationship based on trust and an open dialogue with civil society, the Health Data Hub promoted improved access to information in 2021, particularly concerning the exercise of each patient's rights. Its communication actions were guided by citizens' needs identified through working groups, including user representatives, and citizen consultations, managed by the Citizen Relations Department. To this end, a partnership strategy with patients' associations was established, as for the VivreCovid19 study conducted by France Assos Santé in collaboration with the HDH.

Expanded strategic partnerships. To carve out its own place in the health data ecosystem, the HDH has forged patient-oriented partnerships with several major health players. With a view to rolling out a broader range of services, some twenty partners have been identified among key players in the ecosystem of registers, national databases, cohorts, health establishments, health agencies, etc. Some of these partnerships will lay the foundations for future links between the HDH and local or thematic initiatives. Examples include collaborations with the CNAM, INSERM, AP-HP, Unicancer, Datalab Normandie and the Groupement de coopération sanitaire des hôpitaux du Grand Ouest.

An innovative and secure technological infrastructure. Development work on the Health Data Hub's technological platform began in 2019, and it was brought into service in 2020. This platform poses technological challenges as it must meet stringent security requirements. It must also provide a software package capable of meeting the needs of innovative projects that in some cases have previously been impossible to conduct in France. In this context, reversibility studies have been carried out in collaboration with our partners such as DINUM (Interministerial Digital Technology Directorate). Cybersecurity governance involves regular audits, the last of which was conducted by the ANSSI (French National Agency for the Security of Information Systems) in the summer, after an audit conducted in February 2021.

The internal structure of the HDH. The HDH’s structure has been developing rapidly to meet the challenges set out in the Strategic Roadmap for 2020-2022. This process continued throughout 2021 and included a significant consolidation of the teams.

HDH HIGHLIGHTS AND KEY FIGURES

61 INNOVATIVE PROJECTS

41 PARTNER PROJECTS

14 PROJECTS RELATED TO THE EPIDEMIC

6 PROJECTS FROM THE CALL FOR EXPRESSIONS OF INTEREST (CEI) ORGANISED WITH DATALAB NORMANDIE

4 PROJECTS FROM THE BOAS CEI

6 NEW DATA CHALLENGES

3 FUTURE PROJECTS IN THE UNIBASE PROGRAMME WITH UNICANCER

… MAKING PROGRESS

30 CESREES-APPROVED PROJECTS OUT OF 44 ELIGIBLE PROJECTS

16 PROJECTS AUTHORISED OUT OF 44 ELIGIBLE PROJECTS

3 PROJECTS COMPLETED OR BEING FINALISED

4 ACADEMIC PAPERS PUBLISHED OR PENDING PUBLICATION

7 TOOLS ALREADY AVAILABLE ON THE TECHNOLOGICAL PLATFORM

18 DATABASES CONSIDERED FOR INCLUSION IN THE FIRST VERSION OF THE CATALOGUE

INTERNATIONAL INFLUENCE

7

INTERNATIONAL MEMBERS OF THE SCIENTIFIC ADVISORY BOARD

15

PARTNERS IN THE EHDS APPLICANT CONSORTIUM

3 **PARTICIPATION IN EUROPEAN CALLS FOR PROJECTS**

3 **EVENTS ORGANISED BY THE HDH:**

700

participants in the "AI and Medicine: Promises and Limitations" event, co-organised with MIT in May 2021

350

participants in the 3rd edition of the "Health Data and Collective Intelligence" conference organised by the HDH with 17 speakers in December 2021

428

participants in the "AI4HEALTH" Winter School in January 2022

A RESOLUTELY CITIZEN-ORIENTED PROJECT

6 **PARTNERSHIPS WITH ASSOCIATIONS**

1 **EUROPEAN E-CONSULTATION**

2

REVIEW OF THE HEALTH DATA HUB'S STRATEGIC ROAD MAP

FOUR KEY PRIORITIES OUTLINED IN THE 2020 - 2022 STRATEGIC ROADMAP

In January 2020, the HDH published its strategic roadmap, based on four major strategic priorities:

Opening up health data assets and improving their quality;

Developing an efficient service offering for health data users;

Positioning France as a leader in the secondary use of health data in the European and international ecosystem;

Ensuring the active participation of civil society in health data uses.

PRIORITY 1

OPENING UP HEALTH DATA ASSETS AND IMPROVING THEIR QUALITY

One of the HDH's missions is to make a catalogue of databases available to project leaders. This catalogue consists of a collection of pseudonymised databases that the HDH is authorised to make available. Discussions are currently underway with around 20 potential partners including Santé Publique France, AP-HP, INSERM, the Université de Bordeaux and Unicancer.

Action 1.1 – Adding priority datasets to the catalogue

The Law of 24 July 2019 on the organisation and transformation of the French health system extended the National Health Data System (Système National des Données de Santé – SNDS). In addition to medico-administrative data, the SNDS now includes all health data associated with public funding, i.e. data from registers, research cohorts and hospital data warehouses, etc. The large medico-administrative database, historically encompassing the SNIIRAM (National Health Insurance Information System), the PMSI (Medicalisation of Information Systems Programme) and the CépiDc (Epidemiological Centre for the Medical Causes of Death Database), is now called the "main SNDS database".

Although the HDH is legally responsible for compiling SNDS data, organising it and making it available, this does not mean that this data is or will be collected in a single file or even on a platform.

The HDH is empowered to make this information available to projects, either on an ad hoc basis – the specific data required by a project is provided and then deleted – or on a more permanent basis in order to pool the efforts involved in transferring the data required for several projects.

The source databases are never modified; they are simply copied, in compliance with the regulations. When these copies are made, directly identifiable personal information is removed and the individuals concerned are informed.

This copying of databases is intended to serve a large number of projects, and is highly regulated. The databases concerned are listed in a decree issued by the Ministry for Solidarity and Health after consultation with the French Data Protection Authority (CNIL).

This is of interest because transferring this data requires both the administrator of the database and the HDH – the recipient of the data – to possess the necessary resources. Coordinating their efforts to copy this data on a single occasion, with updates, enables the actors to pool their efforts, either by putting in place the technical transfer procedures and arrangements for informing the persons concerned, or by enriching the database with other sources on the HDH platform.

The other benefit of this approach for the originators of the data is the ability to delegate the following processes to the HDH: contractualisation with users, the preparation, cross-referencing and provision of the data in dedicated and highly secured IT environments per project, and maintenance of the environments.

However, not all databases can be included in this scheme. They are therefore selected, and the final collection forms what is referred to as the HDH "catalogue", which is therefore being gradually compiled, in conjunction with the actors who created or administer the databases concerned, under the aegis of the Strategic Committee of the National Health Data System, created in 2021 by the SNDS decree.

18

PRIORITY DATABASES FOR THE FIRST VERSION OF THE CATALOGUE

Publication of the decree – another step in the implementation of the 2019 Law

The public authorities defined the governance system for the SNDS in the SNDS decree of 29 June 2021, making the HDH along with the French National Health Insurance Fund jointly responsible for processing its core database. This decree also appointed the HDH as the data controller for the SNDS catalogue.

The decree also modifies the list of organisations, establishments and departments benefiting from permanent access to the SNDS data, due to their public-service remits. It sets out the conditions for informing individuals and arranges for the application of the rights provided for by the GDPR in the context of the SNDS. For example, the right to object does not apply either to the creation of the main SNDS database or to the provision of this data to bodies with permanent access. Nevertheless, individuals may contact the Director of the HDH or the Director of the compulsory health insurance managing body to which this data belongs in order to exercise their right of access, right to rectification or right to restrict the processing of data in the core database.

In order to structure and improve access to the data in the SNDS, the decree established the distinction between the "core database" and the "catalogue". It also led to the creation of a Strategic Committee, tasked in particular with identifying existing databases that should be included in the catalogue, in order to assist the minister with defining the general guidelines for the development of the SNDS.

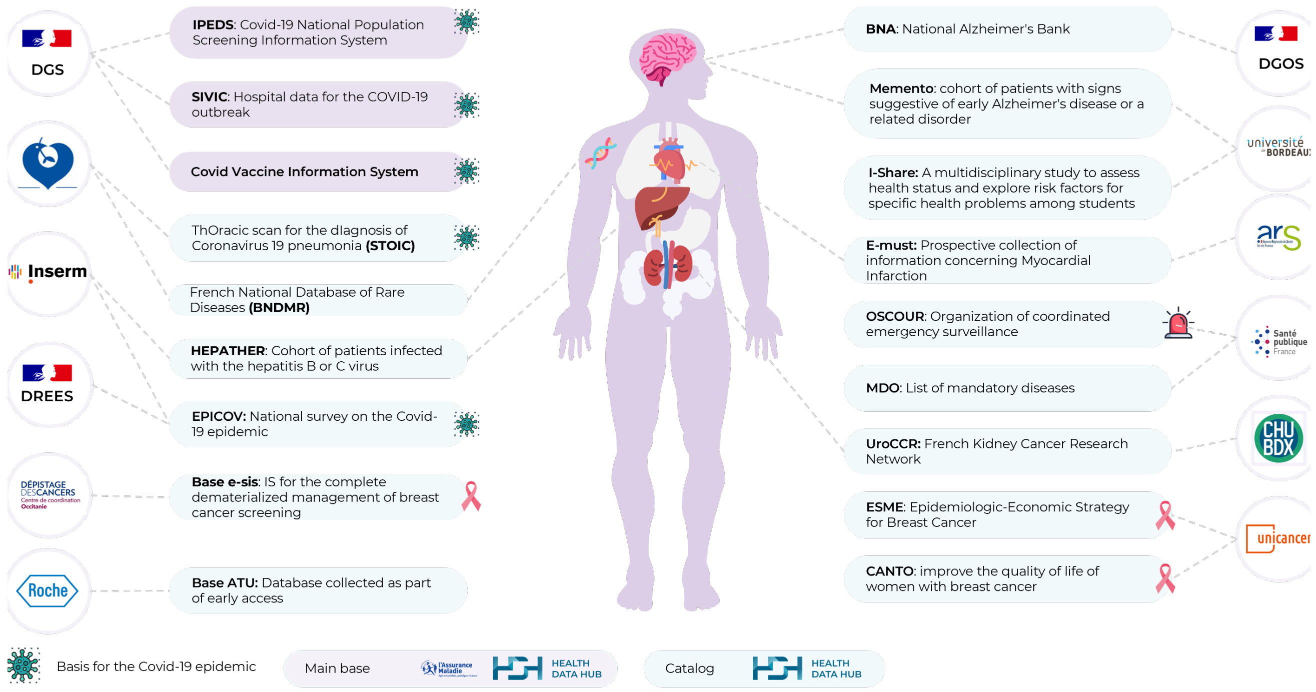
Indeed, the list of databases included in the catalogue must be published in an order of the Minister of Health. This order must also specify the projected flows in the main base, in addition to the historical flows. A number of information systems that include data on the epidemic may therefore join the main database in the future. This decree is still being examined by the French Data Protection Authority (CNIL), and discussions are ongoing between the CNIL, the Directorate for Research, Studies, Evaluation and Statistics (DREES), and data controllers at the HDH and the French National Insurance Fund (CNAM).

Once this legislation has been published, the system will be supplemented with the submission to the CNIL of a request for authorisation to use the HDH's technological environment, which is already used for one-off projects, for the purpose of hosting these different data sources on a permanent basis.

In practice, this means that the HDH teams may also be able to produce extractions from the main database for use by authorised project leaders. This capability will supplement the service offering of the French National Health Insurance Fund, and will facilitate the provision of this data in the HDH's technological environment when desired by project leaders. The same will apply to the other databases in the catalogue since, pursuant to the SNDS decree, the HDH is now responsible for storing and ensuring the availability of the data in the main database and all the databases in the catalogue, as well as for additional pseudonymisation operations for the data it makes available, after the CNAM has carried out an initial pseudonymisation to enable potential matching between the databases and with the copy of the main database of which the HDH is a recipient.

Pending the publication of the final legislation, the HDH and the CNAM have been working actively for several months to roll out this new activity, by focusing on the technical aspects in preparation for the transfer, but also on upskilling the HDH's engineers. In the future, these specialists will be required to carry out expertise-intensive activities which are currently carried out mainly by the French National Health Insurance Fund .

A list of prospective databases for the first version of the catalogue



IPEDS: Covid-19 National Population Screening Information System

SIVIC: Hospital data for the COVID-19 outbreak

Covid Vaccine Information System

ThOracic scan for the dIagnosis of Coronavirus 19 pneumonia **(STOIC)**

French National Database of Rare Diseases **(BNDMR)**

HEPATHER: Cohort of patients infected with the hepatitis B or C virus

EPICOV: National survey on the Covid-19 epidemic

Base e-sis: IS for the complete dematerialized management of breast cancer screening

Base ATU: Database collected as part of early access

BNA: National Alzheimer's Bank

Memento: cohort of patients with signs suggestive of early Alzheimer's disease or a related disorder

I-Share: A multidisciplinary study to assess health status and explore risk factors for specific health problems among students

E-must: Prospective collection of information concerning Myocardial Infarction

OSCOUR: Organization of coordinated emergency surveillance

MDO: List of mandatory diseases

UroCCR: French Kidney Cancer Research Network

ESME: Epidemiologic-Economic Strategy for Breast Cancer

CANTO: improve the quality of life of women with breast cancer

Basis for the Covid-19 epidemic

Main base — l'Assurance Maladie — HEALTH DATA HUB

Catalog — HEALTH DATA HUB

2022 Roadmap

First list of potential bases for the HDH catalog

Commitments to data controllers

In 2020, the HDH made several commitments on data sharing to collectors of this information, i.e. data controllers.

- To support the collection, standardisation and documentation of data. In addition to contributing financially, it can provide expertise to help them prepare and transfer this data.
- To provide a secure hosting service and assistance with GDPR compliance.
- To promote their activities and expertise. To this end, it raises the profile of research databases, thereby increasing the number of citations mentioning their re-use, and notifies data controllers whenever results benefiting from their data are published. Finally, it promotes, or even supports, partnerships between data controllers and users, and enables them to charge a fee for their services.

Our service offering designed for data controllers

Service contract and terms

OFFERING	PRINCIPLE	STATUS
Standard agreement	A standard, ready-to-use agreement specifies the main principles of the partnership between the Health Data Hub and the data controller for the inclusion of databases in the main catalogue. Now finalised, this standard agreement is shared with numerous data controllers. However, the conditions for the economic exploitation of access to data are still being examined.	✓ **Finalised**
Technical procedure for sharing	The HDH provides a secure transfer system on three levels: creation of a virtual private network (VPN), provision of a secure file transfer protocol (SFTP), and encryption of the data transmitted by the data controller. The transfer channel is currently being put in place with several partners.	✓ **Finalised**
Open science	The HDH aims to promote the sharing of knowledge and resources relating to the main SNDS database and the other databases under open licence. This sharing is initiated by different contributors and users.	✓ **Finalised**
Promotion of the database via the metadata catalogue	The Health Data Hub undertakes to promotes the data controller's staff, governance and activities, and sets out to raise their profile within the scientific community. For this purpose, it distributes an identity form associated with the database. A more detailed metadata catalogue, which can be progressively enriched, is currently being rolled out.	✓ **Finalised**
Pricing	The possibility of charging a fee is being considered in order to offset some of the collection costs, but this must be carried out in compliance with complex regulations. An initial detailed analysis of the different legislation likely to interfere with the implementation of a fee structure was finalised in 2021. A proposed model will be developed in 2022, in collaboration with the HDH's administrative bodies.	**In progress**

Regulatory support

OFFERING	PRINCIPLE	STATUS
Database compliance	The Health Data Hub helps any actors who wish to verify the compliance of the database to do so by offering them a self-assessment kit.	⚙ In progress
Patient information	Standard information statements concerning the transmission of data to the Health Data Hub (catalogue and matching with the main SNDS database) have been finalised and shared with numerous data controllers. Three profiles were identified in order to adapt the messages. These statements should also be adapted to the data controller's existing communication channels in order to facilitate their dissemination. At the same time, the Health Data Hub will be providing a second level of information on its website.	✓ Finalised
"Right to object" procedure	The Health Data Hub guarantees the exercise of individuals' rights, in particular by publishing the Public Directory on its website,and the procedure explaining how individuals can exercise their rights. An objection procedure has been finalised and shared with the data controllers concerned. A set of specifications is being finalised to produce a form that will improve the practical procedures for exercising rights.	✓ Finalised
Contacts with CESREES	Data controllers who populate the catalogue liaise with the French Scientific and Ethical Committee for Research, Studies and Evaluations in the Health Sector (CESREES). Their expertise contributes to the assessment of the feasibility, methodology, relevance or ethical nature of a study that uses the data they have generated. The HDH serves as the secretariat of this committee and organises these interactions.	⚙ In progress

Spotlight on COVID

The COVID-19 epidemic has demonstrated the importance of sharing health data. Fully mobilised at the outbreak in the spring of 2020, the HDH remained actively involved in 2021 by automatically accepting any project related to the epidemic, even though the HDH's services were still in the test phase.

The government order on the state of health emergency, which came into force on 21 April 2020, supplementing the order of 23 March 2020, enabled the HDH to collect and provide access to data relating to the epidemic for research purposes, and was incorporated into ordinary law on 31 May 2021. It will remain in force until the order establishing the bases for the HDH catalogue is published.

Throughout 2021, four new projects related to the epidemic were submitted to the HDH, bringing the number of pandemic-related projects currently supported to 14.

These projects are varied and enable different aspects of the phenomenon to be studied. They set out to improve forecasting of the progression of the epidemic, analyse social inequalities in health and study the care pathways of certain populations with co-morbidities, which have been hit particularly hard. The time taken to obtain the regulatory approvals and the time required to complete these projects meant that they were not completed in 2021, but some are expected to start producing results in H1 2022. Indeed, all of them have been approved by the French Scientific and Ethical Committee for Research, Studies and Evaluations in the Health Sector (CESREES) and more than a third have been authorised by the CNIL.

Action 1.2 – Establishing an automated national procedure to facilitate data selection, addition to the catalogue, and matching

Implementation of the service offering to facilitate the management of matching

As the data are dispersed throughout the country, it is essential to be able to cross-reference them in order to facilitate their mutual enrichment or to obtain the critical mass required for certain projects.

Ideally, in order to cross-reference these data, a common pivot variable is required. In the world of health data, this variable can be a person's social security number, where available. However, as this identifier is subject to a strict regulatory framework, it must be transformed in order to protect it and prevent the identification of the people in question, while enabling the performance of research activities.

However, to retain its pivotal role, this variable must always be processed in an identical manner. A transformation procedure is therefore used, involving different entities belonging to the main national health insurance system, which ensures that the same actors do not have access to the initial social security number and the final result of its transformation.

Mobilising this procedure remains complex and a number of actions must be carried out manually. A substantial increase in the number of projects and the prospect of growing demands for matching means that this procedure will need to be implemented on a much larger scale. This is why the HDH and the CNAM are working on a tool, called a "concentrator", that is designed to automate certain steps and make this procedure more accessible to its many beneficiaries. There are plans to extend its functionalities to the automation of cross-referencing involving the reconstitution of social security numbers by the CNAV (National Old-Age Insurance Fund).

In addition to facilitating matching, it should be noted that this tool will also enable the manipulation of social security numbers in complete security, to uphold the rights of the individuals concerned and enable anyone to object to the re-use of personal data for research purposes. This feature will be implemented in the form of an interface accessible to all citizens.

A first set of specifications for this tool was designed in 2021, and their practical implementation is scheduled for 2022.

Action 1.3 – Creating a "core cohort" compiling precisely located clinical, medical and administrative data for a representative sample of the population

Several existing databases have been selected for sharing in the Health Data Hub catalogue. Nevertheless, the links between healthcare institutions, their health data warehouses and the Health Data Hub also hold enormous potential for the creation of thematic, multi-centric databases, and for the transmission of data on demand. Several actions have therefore been carried out in 2021 to explore these possibilities.

A partnership with Unicancer concerning the UNIBASE programme

With more than 380,000 new cases per year, cancer is the leading cause of death in men and the second cause of death in women in France. To accelerate research and improve these patients' survival, it has become imperative to link the available data and exploit it in a secure technological environment enabling the use of artificial intelligence.

To this end, in July 2021 the HDH signed a partnership with Unicancer, a major player in innovative approaches to fighting cancer. The UNIBASE programme is at the heart of this partnership, which aims to create a collection of oncology reference databases over a three-year period, based on data from French Cancer Control Centres (Centres de Lutte contre le Cancer – CLCC) located nationwide. This data will be made available to enable stakeholders in the research sector to roll out collaborative, multi-centric studies and tackle scientific challenges that would be impossible to overcome without access to massive real-world data. The centralisation of data, the cross-referencing of multiple sources and the combination of clinical, imaging and omic data make UNIBASE a unique programme in oncology.

The HDH and Unicancer launched its concrete implementation on 25 November 2021 by issuing a call for expressions of interest, open to the ten French cancer control centres (CLCCs) equipped with data warehouses based on Consore, the search engine for big data in oncology. Three pilot projects, consistent with the 2021-2030 Cancer Plan, will be selected and rolled out on the HDH's technological platform from 2022 onwards, with support and funding from the HDH.

Based on this experience, Unicancer and the HDH are subsequently planning to open up UNIBASE to CLCCs not using Consore, as well as to other hospital facilities. Work will be carried out to ensure that data is interoperable and modelled in accordance with international standards.

3 **PILOT PROJECTS WILL BE SELECTED UNDER THE UNIBASE PROGRAMME**

A partnership programme for healthcare institutions aiming to carry out multi-centric research projects by pooling their data

In 2021, the HDH launched a partnership programme open to any healthcare institutions interested in consolidating their data strategy and collaborating with the HDH in order to explore solutions to enable the pooling and enrichment of their hospital data. In concrete terms, this involves carrying out research projects using health databases shared by various hospitals. The HDH will assist them with the coordination of these projects, provide financial support, and will give them access to its technological platform to enable the cross-referencing of the consolidated data with data from the main SNDS database and the implementation of the identified type of use. The programme also aims to improve these actors' networking capacities (e.g. their ability to collaborate on a collective response to strategic supranational opportunities) and to develop their expertise by sharing their experience.

17 **NUMBER OF INSTITUTIONS WITH WHICH THE HDH WORKS**

The Health Data Hub wins a call for tenders from the European Medicines Agency

At the end of 2021, the European Medicines Agency selected the application submitted by the HDH and its partners in response to a call for tenders on real-world data. Objective: to gain access to individual hospital databases from EU countries at the patient level and in OMOP-CDM format.

To meet this challenge, the Health Data Hub has joined forces with four health institutions: Hospices Civils de Lyon (HCL), the Centre Léon Bérard, Nancy Regional University Hospital and the Hôpital Saint-Joseph, as well as the Bordeaux Pharmaco-Epidemiology platform, the French Digital Health Agency (ANS) and Capgemini. Together, they will consolidate a hospital clinical database, enriched with medico-administrative data from the main SNDS database and transformed into a common international model to ensure its interoperability. The European Medicines Agency will be able to use this data to investigate various issues related to the side effects of medicines. For all partners, this project is an opportunity to develop common working methods that will allow the experiment to be repeated whenever relevant, in response to national or international issues.

The HDH is a member of the Non-Hospital Care Observatory (Observatoire de médecine de ville) consortium.

The HDH is a member of the consortium led by the National Board of Teaching General Practitioners (Collège national des généralistes enseignants – CNGE), which is in the process of developing a "non-hospital care data warehouse", i.e. covering private health professionals and medical care provided outside of hospitals or medico-social institutions. This consortium includes the start-ups Loamics and MyDataModels, as well as Rouen University Hospital, Université Côte-d'Azur and the Université de Normandie, with the support of the National Health Insurance Fund (CNAM) and the French Ministry for Solidarity and Health.

In 2021, it accomplished the first stage of its call for expressions of interest for the "Digital Health" acceleration strategy.

This project is of major interest because it will:

- Provide a complete overview of care pathways, including upstream and downstream of any hospital stays, by combining data from consultations in the non-hospital care sector with SNDS data;
- Detect early warning signals for the resurgence of an epidemic by monitoring trends in diagnoses from nationwide consultations;
- Identify risk factors more precisely, by improving knowledge of the symptomatology of diseases, including, as a priority, COVID-19.

Although the need for such an observatory has long been recognised, the health crisis has accelerated this project. At a time when the role of non-hospital medical care needs to be defined alongside and in addition to the emergency services, the large-scale collection of data from the non-medical sector provides valuable support for the management of epidemics such as COVID-19. However, this data was collected in more than 20 different electronic medical record solutions used by doctors, and is currently neither harmonised nor consolidated at the national level. The project to create a Non-Hospital Care Observatory aims to create a data warehouse linked to the SNDS via a circuit designed to transfer data from the main software solutions used by general practitioners.

By including the HDH in this consortium, the project leaders, from the outset, wanted to provide for the sharing of data for general-interest purposes via the HDH data catalogue.

Action 1.4 – Promoting producers that contribute to the data-sharing movement

To promote producers who participated in the data-sharing process in 2021, the HDH, for example:

DEVELOPED A PROTOTYPE METADATA CATALOGUE TO HIGHLIGHT SHARED DATABASES

CARRIED OUT A LEGAL ANALYSIS TO GOVERN THE IMPLEMENTATION OF THE FEE AND DEVELOP A CHARGING SYSTEM

ORGANISED A NEW EDITION OF ITS ANNUAL "HEALTH DATA AND COLLECTIVE INTELLIGENCE" CONFERENCE

A metadata catalogue

To facilitate access to health data and use it more intensively in research projects, "merely" providing access to this data is not sufficient. The content of the available databases must also be identified and precisely documented, in order to accelerate the users' search for data that can answer their research questions. This is why the Health Data Hub collects and provides online access to the ***metadata*** contained in each of its catalogue's databases.

In concrete terms, the consultation of metadata should speed up access to data, optimising interactions between researchers, producers and the HDH. It also provides researchers with the producers' contact details, thereby promoting interactions and collaborations between scientists. Finally, opening up access to metadata reflects the HDH's desire for transparency, as it provides the maximum amount of details about the data collected for research purposes, without disclosing any identifying information.

Technically, this ***metadata catalogue*** has three levels of information (Base, Tables, Variables), and has a search engine to identify the HDH databases that correspond most closely to the user's query. It then enables users to explore the content, from the general description of a database, its scientific context, and the methods used to collect and provide access to the data, to the technical and statistical description of each variable it contains.

The solution that has been adopted draws on the specialised knowledge of numerous experts, and its implementation is based on the CKAN open-source data management system, for which the HDH has written an extension available in a public GitLab repository. It forms part of an international approach to the standard description of health data and will eventually guarantee the best means of accessing and interacting with other health data platforms such as the European Health Data Space (EHDS).

A fee for private actors seeking direct access to catalogue data

To boost the sharing of health data, in 2021, the HDH examined the case for introducing a fee for private actors seeking direct access to the data in its catalogue. This review involved a legal analysis phase to identify the legal and regulatory framework for a possible fee system. Several data controllers were asked about their pricing practices. They considered that these practices do not enable them to guarantee the long-term functioning of the different public health data bases and infrastructures. In line with this observation, the HDH developed a case for increasing the public funding of strategic data infrastructures in health care institutions. The second phase will begin in 2022 and is aimed at building and implementing the future fee-charging system for access to HDH catalogue data.

Focus on the third edition of the "Health Data and Collective Intelligence" Conference

Bolstered by the success of the two previous editions, the HDH held a new edition of its annual "Health Data and Collective Intelligence" Conference ("Données de santé et intelligence collective") on 15 December 2021. An event co-organised with the Ministerial Delegation on Digital Health and the "Grand Défi" (Major Challenge) programme entitled "Improving medical diagnoses with artificial intelligence", on the PariSanté Campus site. More than 350 participants attended online and interacted with some 20 speakers. On the agenda: the benefits of health data for research, public health and the management of the health system, but also the potential benefits in terms of attractiveness and competitiveness that reusing this data can offer France.

350

PEOPLE PARTICIPATED IN THE "HEALTH DATA AND COLLECTIVE INTELLIGENCE" CONFERENCE

PRIORITY 2
ENHANCING DATA USE BY OFFERING ATTRACTIVE SERVICES

Before health data can be used for research purposes, three stages must be accomplished: the regulatory stage of the application for access, the process of actually accessing the data, and the implementation of the project. The HDH and its partners are involved in making the improvements required in each of these stages.

Action 2.1 – Establishing simplified procedures and a large-scale access process

The HDH accompanies project leaders throughout their database-access applications, with indispensable support provided by its partners' teams, especially the database administrators.

366

FILES REQUIRING CNIL AUTHORISATION SUBMITTED TO THE HDH IN 2021

How does CESREES examine applications to access health data?

Created in 2020, the French Scientific and Ethical Committee for Research, Studies and Evaluations in the Health Sector (CESREES), for which the HDH serves as the Secretariat, gives its opinion on each health-data-access application. Its opinions concern:

- the purpose and methodology of the research,
- the need to use personal health data,
- the ethical relevance,
- the scientific quality of the project,
- and, where appropriate, the public interest served by the project.

279

CASES APPRAISED BY CESREES IN 2021

The HDH's first contribution to the CNIL's simplified procedure

On 26 October 2021, the French Data Protection Authority (CNIL) published a "standard relating to the processing of personal data carried out in order to create data warehouses in the health sector".

This simplified procedure enables any data controller who meets the conditions set out in this standard to establish a health data warehouse, after submitting a compliance commitment to the CNIL. A health data warehouse is defined by the Commission as a massive collection of data, compiled over an extended period of time, for the purpose of research, studies and evaluations in the health field.

Prior to this publication and in accordance with Article 66 II of the French Data Protection Law, the CNIL had asked the HDH to give its opinion on this standard. In response, the HDH consulted representatives of its Board of Directors on a legal analysis of the draft text, in order to gather everyone's observations. This contribution was forwarded to the CNIL and enabled the authority to adjust the draft text, taking account of the actors' needs.

Throughout 2022, the HDH will have opportunities to contribute to new simplified health-data-access procedures in France.

Continuous improvement of the service offering in 2021

To give projects every chance of obtaining a favourable opinion from CESREES and of subsequently being authorised by the CNIL, a sound understanding of the procedure, its various stages, and the content of the file to be submitted are required. These aspects may seem complex, especially to new players on the scene. This is why the Health Data Hub, the CNAM and CESREES have designed a "starter kit" for project leaders. This educational kit lists points requiring vigilance, provides advice and examples for correctly completing the application, and also includes a complete dummy protocol and a check-list to enable project leaders to identify the expected requirements in concrete situations.

A second educational guide has also been published in conjunction with the CNAM in order to help any applicants for access to SNDS data to complete their expression of needs, which is essential to the precise definition of the scope of the data to be extracted.

A third tool, published in 2021, was added to offer project leaders guidance with diverse health data access procedures. This guide enables these actors to characterise their projects and gives them key information to help them choose the right regulatory procedure to implement.

In addition to these tools, the HDH also set up two training courses in 2021 to help applicants for access to health data.

- Training on SNDS data and how to access it. Run in collaboration with the French National Health Insurance Fund, this three-hour session is aimed at any project leaders who wish to use these data but who may not always be aware of their content, their means of access or the associated regulatory procedures.
- A presentation of the procedures for accessing personal health data. This one-hour

video-conference session is aimed at any project leaders who are seeking to identify the procedures best suited to their projects. It describes the key elements for characterising research, whether or not it involves human subjects, and sets out the regulatory access procedures.

6 TRAINING COURSES AVAILABLE ON 7 OCTOBER 2021

Action 2.2 – Improving and maintaining the HDH'S state-of-the-art technological solutions

A secure "analysis area" on a technological platform

The HDH supports researchers, entrepreneurs and patient associations wishing to carry out and accelerate their public-interest projects, by providing them with a secure "analysis area" on a technological platform, where they can receive, store and process and visualise the health data required for the projects without being able to export them.

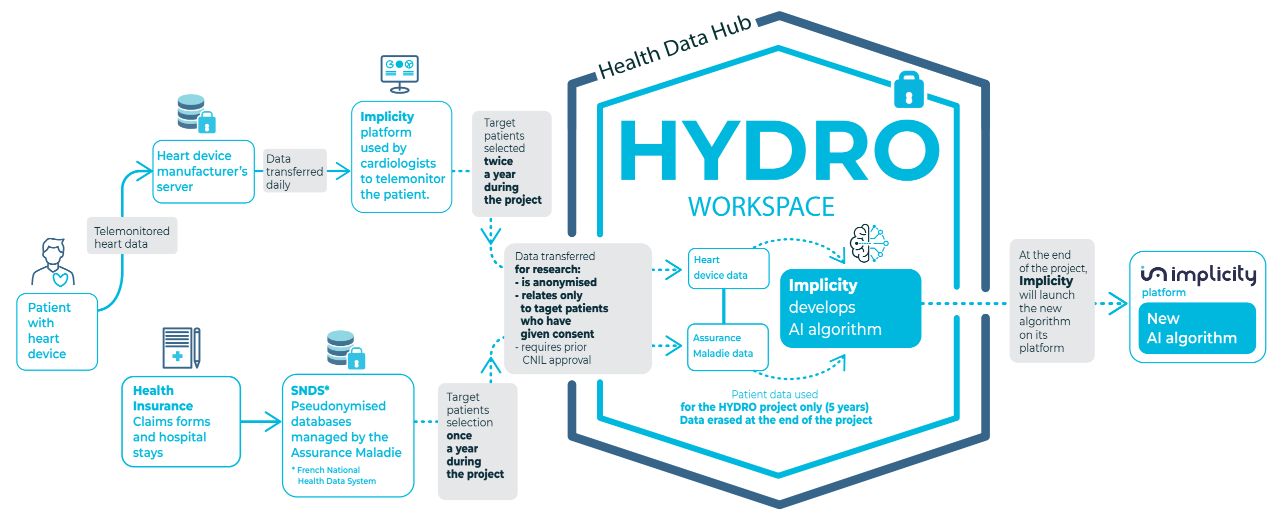
HYDRO by implicity

A research project on artificial intelligence (AI) led by **Implicity**, in partnership with the **Health Data Hub,** and supported by the **Assurance Maladie**

1ST PILOT PROJECT BY THE HEALTH DATA HUB

An AI algorithm to predict episodes of heart failure and reduce risks of hospitalisation for patients fitted with cardiac implants.

PATIENT DATA PATHWAYS WITHIN THE PROJECT



There is no external access to this project space, i.e. it is specific to the project in question, and data cannot be transferred to or from local workstations. Based on an "elastic" infrastructure, it offers variable-capacity calculation resources tailored to the development of the most innovative uses.

The security context for the Health Data Hub platform was defined in collaboration with the ANSSI (French National Agency for the Security of Information Systems) and the Senior Defence and Security Official for the Ministries of Social Affairs and Health. The security of the technological platform is also constantly tested and reviewed in technical audits performed by independent service providers certified by ANSSI. Finally, ANSSI and the CNIL participate in the approval of each new version of the technical platform.

In 2021, a PASSI audit of the technological platform was carried out prior to its approval review meeting on 30 November 2021. An ANSSI audit had been successfully conducted earlier this year. As part of this review, the HDH also updated its risk analysis using the EBIOS RM methodology. This revision took into account new threats and identified emerging risks. As part of the continuous improvement process, a specific risk- management plan was put in place following these reviews.

To address key security issues, the HDH needs to work hand in hand with its partners. For example, it has put in place an ANSSI-certified security procedure for data transfers by data controllers and a best-practice guide to pseudonymisation shared with the CNIL. The HDH has also launched discussions with the French Institute for Research in Science and Automation (INRIA) about a partnership on the analysis of suspicious queries.

A constantly evolving technological offering

Using innovative analytical approaches to process health data requires the resolution of many technological challenges. This is why the HDH provides a constantly evolving software package for all project leaders in a secure, state-of-the-art environment.

To facilitate the performance of their studies, it offers them a "project space" in which they can receive, store, collaboratively process and visualise the health data required for their projects, in complete security.

The HDH is attentive to its users' needs and in 2021, it implemented a product roadmap designed to prioritise their needs in terms of software solutions.

The HDH currently offers the following software

The JupyterLab and RStudio Integrated Development Environments (IDEs), which enable users to work with the Python and R programming languages, respectively.

The PySpark and Sparklyr modules, enabling the processing of large volumes of data

The Superset data visualisation tool

The PostgreSQL and MySQL database management systems

The Cytomine image viewing and annotation tool

The Git and Gitlab collaboration tools

Focus on reversibility

Pursuit of the reversibility action plan

Exclusively administered by HDH staff with specific roles and responsibilities, the technological platform uses industrial components chosen on grounds of their performance, security and elasticity.

State-of-the-art software solutions will enable the development of codes and algorithms (R, python, spark). A marketplace will eventually be launched to facilitate the sharing of relevant tools with the community.

The platform must meet the following key requirements:
> Security: ensuring the protection of highly sensitive health data stored on the platform.
> Functionalities and performance: using data-science-based approaches to meet the needs of public-interest project leaders.
> Costs and deadlines: rapidly implementing a solution to meet health research challenges in a context of international competition.

Designed from the outset to be reversible in the absence of sovereign solutions meeting these three objectives, the technological platform is based on certain solutions that do not correspond to the national "Trusted Cloud" strategy. However, the efforts to identify such a solution are continuing with a third update of the market study that sets out to compare cloud solutions likely to address the HDH's key issues, while conforming to the new national requirements. This study is more extensive than the previous market studies and was largely carried out in 2021. It is currently being finalised and will be enriched with new emerging solutions, including hybrid solutions.

At the same time, the production of platform documentation in Terraform format is continuing, in preparation for migration to the target solution. In addition, the reduction of dependency on the current suppliers is continuing, where possible.

Finally, in response to the growing challenge of digital sovereignty, the HDH is closely monitoring the discussions taking place within various European bodies.

Action 2.3 – Developing a service offering to enable efficient data use

Open leadership of the science ecosystem

Although the HDH is clearly committed to enabling faster access to health data, there are no guarantees that the data obtained will produce convincing, reliable and relevant results. To optimise the exploitation of this data, considerable work will be required in order to understand its organisational structure, its collection procedures, and its biases and limitations. To this end, the HDH provides free access to information, documentation and training for all. It actively implements the open science policy promoted by the Ministry of Higher Education, Research and Innovation.

In practice, since 2019, the Health Data Hub and its partners (DREES, Santé publique France, CNAM, ANSM, ARS, ATIH, ReDSiam, etc.) have been providing open, collaborative documentation related to the main SNDS database, whose content will be supplemented in 2022 with new entries and an introductory guide to the SNDS. This documentation is accompanied by a series of tools to familiarise users with the handling of data, including instructional notebooks, a synthetic data generator, a dictionary of tables and variables, and data handling programs. These tools, consulted about 500 times a day, whose source codes are open to all in Gitlab repositories, are informed by the community and maintained by the HDH.

New projects and tools initiated in 2021 will be added from 2022. The Open Library of Algorithms in Health (Bibliothèque Ouverte d’Algorithmes en Santé – BOAS) project, carried out in collaboration with the ReDSiam network, is one of these initiatives. By supporting and assisting with the development and validation of targeting algorithms in the SNDS, the HDH will enable everyone to access implementations and documentation that are essential to the reliable reuse of these algorithms, which will facilitate the creation of study populations. A concrete manifestation of this project is the provision of support for four initial algorithm validation projects, selected from responses to a call of expressions of interest, which will continue in 2022. The HDH has also mapped the actors involved in research on the SNDS, compiling a dashboard of information associated with over 900 publications. Finally, the HDH's work on data standardisation and the use of a common data model opens up new prospects for the sharing of programmes and knowledge.

11 **CONTRIBUTORS TO THE OPEN-SOURCE BASED APPROACH**

1,587 **COMMUNITY MEMBERS FOR 30 MEET-UPS ORGANISED**

6,049 **VISITORS TO SNDS OPEN-ACCESS DOCUMENTATION IN NOVEMBER 2021**

First 4 projects selected on 8 November 2021

REPÈRES, SPF, CHRU BREST

Project to develop and validate an algorithm to identify subtypes of strokes treated in hospitals

BPE, REPÈRES, EDMUS, MAGENDIE NEUROCENTRE

Project to evaluate and improve the diagnostic performance of an algorithm for identifying multiple sclerosis relapses

F-CRIN CRI-IMIDIATE, BNDMR, CEPHÉPI, IPLESP

Project to develop and validate algorithms for rare systemic inflammatory diseases

AHEAD (BPHC, INSERM), CHU BORDEAUX

Project to develop and validate algorithms for orthopaedic medicine, including spinal and shoulder surgery

Action 2.4 - Disseminating the service offering to all actors: research bodies, industrial operators, public-private partnerships, etc.

The Health Data Hub operates as part of an extensive ecosystem, which includes: French institutions – more than 200 researchers are affiliated with INSERM (French National Institute for Health and Medical Research), the CNRS (French National Centre for Scientific Research) and INRIA (French Institute for Research in Science and Automation); major industrial enterprises, including big pharma companies; the start-up ecosystem (notably via France Digitale); and the major European players in health data, including the main national platforms, European research infrastructures, the European Medicines Agency, and various official European working groups focusing on health data.

In 2021, the HDH supported 55 projects. Although it is still too early to attribute medically exploitable advances to these projects, important milestones have been reached that make this a realistic prospect. For example, the Hydro project, coordinated by a start-up called Implicity, has been collectively developed by Implicity, the CNAM and the Health Data Hub, supported by a patients' association. Its objective is to carry out probabilistic cross-referencing of pacemaker data with hospitalization data of sufficient quality to ensure that so that the next stages of the study are not compromised. Reaching this major milestone required several months of commitment and collective intelligence.

Generally speaking, there has been a sharp increase in partnerships with various actors such as Datalab Normandie, with whom a Call for Expressions of Interest was organised at the beginning of 2021. The "Grand Défi" (Major Challenge) programme, entitled "Improving medical diagnoses with artificial intelligence", remained one of our key partners and together in 2021, we coordinated the Data Challenge programme whose results were beneficial to six learned societies. In partnership with Unicancer, the Health Data Hub launched a call for projects at the end of November to lay the foundations for an ambitious project to create a major oncology database.

The HDH is supporting around ten projects in conjunction with the AP-HP, in addition to carrying out strategic research to establish the conditions for successful data sharing and exploitation.

In general, initiatives with consortia of healthcare institutions are being stepped up in order to promote the scaling up of procedures. Partnerships have also been entered into with stakeholders in the world of Open Science, including Redsiam, patient associations that can contribute to data collection or analysis, and private players, not forgetting the partners involved in the creation of the data catalogue: a collection of databases that will be shared within the HDH, with a simplified access procedure governed by a clear and transparent framework defining the conditions for sharing, which is being jointly developed with the stakeholders.

After making significant progress in 2021, these projects are either nearing the end of their study phase or are about to be included on the platform.

PROJECT	LEAD	OBJECTIVE
DeepSarc	**Centre Léon Bérard**	Study the impact of sarcoma treatments using real-world data in order to identify the best treatment regimens
DEEP.TRACK	**CRCDC Occitanie**	Assess the contribution of artificial intelligence to organised breast cancer screening
DSI COVID	**Université Toulouse III**	Describe the link between social precariousness and the risk of infection by COVID-19 and its evolution in France
CoviSAS	**Grenoble - Alpes University**	Identify the prevalence of severe forms of COVID-19 in patients with obstructive sleep apnoea syndrome (OSAS), and the combinations of co-morbidities associated with OSAS that correlate with a higher rate of hospitalisation in intensive care units or death.
CoData Breast Cancer	**ICANS**	Anticipate and prioritise the management of patients with breast disease that was not treated during the health crisis, and analyse the changes in the medical treatment of these patients, their possible losses of opportunity, and the effect of certain changes in practice.
HUGOSHARE	**HUGO Network**	Investigate drug interactions that may cause adverse reactions, based on hospital drug prescriptions
APRIORICS	**CHU de Toulouse**	Develop a morphological description tool for microscopic breast cancer images to improve the understanding of the variability of these cancers.
SEDAAR	**Alphonse de Rothschild Foundation Hospital**	Develop algorithms to create a decision-support and interpretation service for orthoptists and ophthalmologists.
RIAC	**Amiens-Picardy University Hospital**	Develop a set of indicators to monitor COVID-19, detect likely clusters and conduct a retrospective epidemiological study of regulation data in order to provide a better description of the population who called the emergency medical services at the beginning of the crisis but were not always seen at the hospital.
ARAC	**Malakoff Humanis**	Analyse health expenditure in France and medical insurance policyholders' final out-of-pocket expenses.
DAICAP	**AP-HP**	Develop an algorithm that creates standardised MRI reports and predicts the aggressiveness of tumours.

A partnership with Datalab Normandie to improve care through health data

To promote the emergence of projects that improve the quality of care in the Normandy region, the HDH and Datalab Normandie have entered into a partnership, and launched a first call for projects on 29 January 2021. Six projects using health data were selected. Each will receive a minimum grant of €20,000, provided by the Region, mobilising ERDF 2014-2020 funds, and will also receive operational support. Datalab Normandie will facilitate contact between these projects and sources of regional expertise. The HDH will help them develop a better understanding of the regulatory and technical processes.

These are the six projects supported by the HDH: Anthropometer3DNet, H3DMED, OVERCOME, Jumeau Numérique by the VyV Group, PRIMEGE and EDSaNCoh

"Jumeau Numérique" (Digital Twin) – simulating territorial health status

Few tools are available to enable local authorities to simulate the health status of populations in relation to environmental data. Modelling territories by creating "digital twins", as proposed by the VyV group and developed on the basis of health and environmental data, can facilitate the simulation of public policies and their impact prior to their actual implementation.

OVERCOME

The OVERCOME project is led by a consortium involving the Hometrix Health company, the COMETE laboratory at Université de Caen, Caen University Hospital, and the UNA-Pays d'Alençon home-help network. It proposes to collect data from elderly patients who are monitored at home and are at risk of developing progressive cognitive disorders. Combined with data from the main national health insurance system, the aim is to improve the categorisation and prediction of the progression of motoric cognitive risk syndrome, which can develop into Alzheimer's disease.

PRIMEGE

The PRIMEGE project, led by the general medicine association ENSEREVA, aims to create a general medicine database in Haute-Normandie and to develop a connection platform for health professionals, in order to inform the general public and provide feedback to professionals on their practices.

EDSaNCoh

A cohort is a group of patients with common characteristics, who are monitored individually and prospectively, by collecting health data. Rouen University Hospital is seeking to develop EDSaNCoh, a platform to host cohorts created by pooling data from the Rouen University Hospital health data warehouse, L'Assurance Maladie and questionnaires sent to patients.

Anthropometer3DNet

Led by the Henri Becquerel Cancer Control Centre in Rouen, the Anthropometer3DNet project aims to create an online platform for the artificial-intelligence-based automatic analysis of the body composition of patients treated for cancer, based on routine clinical scans. Analysis of the results can provide new insights into major risk factors and improve patients' prognoses.

H3DMED

The François Baclesse Cancer Centre in Caen has specialised for 20 years in personalised curietherapy treatments, particularly for the treatment of gynaecological and skin cancers. The objective of the project is to develop an appropriate AI algorithm to assist physicians with the design of customised curietherapy applicators.

6

PROJECTS SELECTED IN THE DATALAB NORMANDIE CALL FOR PROPOSALS

1

CALL FOR PROPOSALS WITH DATALAB NORMANDIE: "DATA AND ASSOCIATED TECHNOLOGIES APPLIED TO HEALTH"

The partnership with AP-HP revolves around eight pilot projects

The partnership between AP-HP and the HDH, which began in 2020, was consolidated in 2021 by concrete achievements in eight pilot projects and the launch of joint strategic initiatives.

Seven of these eight projects, required CNIL authorisation to operate. Six have already received a favourable opinion from CESREES and the eighth is currently being examined.

PROJECT	COORDINATOR & PARTNERS	OBJECTIVE
EMIR-Algo	**AP-HP**	Conduct a study aimed at developing and validating algorithms for targeting sufferers of various rare autoimmune and systemic auto-inflammatory diseases (SAID) using BNDMR (National Rare Disease Databank) data chained to SNDS data.
BACTHUB	**AP-HP, INSERM**	Establish the link between antibiotic exposure and the occurrence of resistant bacteriemia.
DROMOS	**AP-HP**	Enrich the National Rare Disease Databank with data from the main SNDS database to implement public health indicators.
AI Dream - STOIC	**AP-HP, Therapanacea, GE**	Develop learning algorithms for COVID-19 diagnosis and the assessment of its severity from chest scans.
NHANCE	**AP-HP, INRIA**	Develop and train an AI algorithm capable of detecting all organs in the abdomen and their major injuries.
DIGITAL HPV	**AP-HP, INSERM**	Determine HPV status in upper aerodigestive squamous cell carcinoma and predicting response to therapy.
EPIFRACTAL	**AP-HP**	Evaluate the effectiveness of the Fracture sector and propose a model for automatic patient selection from a cohort based on medical reports.
DAICAP	**AP-HP, INRIA, INCEPTO**	Develop an algorithm capable of creating standardised MRI reports with the prediction of tumoural aggressiveness in prostate cancer cases.

2021 was also notable for AP-HP's commitment to register two databases in the HDH catalogue: the BNDMR database, which will be linked to the SNDS, and the STOIC database. Chaining the BNDMR (National Rare Disease Databank) with the SNDS is a key challenge, as it is likely to increase the research opportunities, thereby ultimately improving the medical treatment of patients suffering from rare diseases, whose conditions may be misdiagnosed for several years.

STOIC is a COVID-19 database containing a large number of chest images of patients suffering from COVID-19 at different stages. Combined with medical annotations, this database extends the range of opportunities for COVID studies.

Other strategic initiatives are under consideration, including a coordinated approach to the OMOP-CDM-formatting of health data, thereby facilitating its international sharing. Discussions are also underway to define the financial and scientific terms and conditions for sharing and exploiting the data that will be included in the HDH catalogue.

The HDH, a founding member of ParíSanté Campus

Announced by the French President on 4 December 2020 and placed under the responsibility of the Ministry for Solidarity and Health, and the Ministry for Higher Education, Research and Innovation, the PariSanté Campus programme aims to create an internationally unprecedented centre of excellence for research, innovation, training and entrepreneurship on the theme of digital health. The HDH is a founding member of PariSanté Campus, along with INSERM, INRIA, Université PSL and the French Digital Health Agency (ANS). By 2028, this world-class campus will be built on the site of the former Val de Grâce Hospital in Paris. In the meantime, a trail-blazing location was inaugurated in the presence of the Prime Minister on 14 December 2021. PariSanté Campus will make a powerful contribution by generating the impetus required to promote meetings between operators and actors in an open location. The HDH's presence on this site will significantly increase its capacity to roll out impactful joint projects that stand out on the international scene. It has already contributed to the construction of a specific offering for start-ups selected by calls for expressions of interest, which will be hosted on the site.

5

FOUNDING MEMBERS OF PARISANTÉ CAMPUS, INCLUDING THE HDH

The HDH becomes a founding member of the Filière Intelligence Artificielle et Cancer (FIAC) association

In 2021, the HDH joined the ten other founding members that make up the FIAC, an association of public and private players involved in a national initiative that seeks to benefit the entire oncological innovation ecosystem and enable the development of innovative therapeutic and diagnostic solutions targeting unanswered questions.

11

MEMBERS OF THE FIAC ASSOCIATION, INCLUDING THE HEALTH DATA HUB

On-demand query service

In 2021, the HDH launched an experiment to design an "on-demand query" tool for actors who are unfamiliar with health data, which will allow them to access data-driven information from the SNDS that is not available in open data format, without any regulatory formalities, ensuring that the items produced and the associated documentation will subsequently be accessible to all.

In 2021, activities to clearly define the scope of this offering were launched in conjunction with various patients' associations such as France Vascularités, Journée Nationale de l'Audition (National Hearing Day Association) and the Association des Porteurs de Dispositifs Electriques Cardiaques (National Cardiac-Device-Wearers' Association). Researchers in the human and social sciences along with journalists and representatives of publications such as the *Quotidien du médecin* also participated in this reflection. These activities will continue in 2022, in preparation for the opening of the service during the year.

PRIORITY 3
POSITIONING FRANCE AS A LEADER IN THE SECONDARY USE OF HEALTH DATA

The HDH provides services aimed at the ecosystem: the initiators of data collection, the users of data, and the general public. To achieve its full potential, the HDH must meet the expectations of an ecosystem that is now operating on a European, or indeed an international, scale.

Action 3.1 – Informing the international scientific community

In January 2021, a Scientific Advisory Board was created to support the HDH's senior management team at the latter's request. Its mission is to inform the HDH's decisions on all scientific aspects that affect its development. Its members meet once a year in plenary session, as well as on an ad hoc basis, according to the HDH's needs. To address issues that may be extremely diverse, 28 personalities from numerous disciplines including medicine, pharmacy, genetics, epidemiology, statistics, health economics, law and ethics sit on this board.

> Since its creation, the Scientific Board has been asked to issue opinions on subjects of interest that are central to the Health Data Hub's development strategy. In particular, it has informed the senior management team's decisions on the orientation of the post-doctoral fellowship funding programme, the construction of an impactful partnership with Unicancer, the scientific value of the databases to be added to the catalogue, and the prospects for the HDH's involvement at the European level.

> ### The post-doctoral fellowship programme
>
> In line with its objective to inform the scientific community about health data issues, the HDH decided to create a post-doctoral fellowship funding programme in 2021. The senior management team sought the Board members' advice on the format of the funding programme and on the relevance and scientific value of pre-identified topics. This funding programme, which is open to all academic research laboratories in France and abroad, provides financial and personal support for the beneficiaries in collaboration with the research laboratory hosting the post-doctoral fellow. To help promote data use for public-interest purposes, beneficiaries of the post-doctoral fellowship programme undertake to provide open access to the results of their research – including any algorithms developed – for the entire scientific community, and to contribute to the dissemination of their work by participating in scientific events.

Three topics have been selected for these first post-doctoral fellowships:

- An international comparative study of the socio-economic, social and ethical impacts of open access to health data,
- Visualisation of care pathways in the main SNDS database.
- Implications of European legislation for the HDH and concrete proposals concerning rights

Cataloguing

The aim of the catalogue is to make databases of interest more easily accessible for health research purposes by replicating them on the HDH technological platform. In its first version, the catalogue will include a list of databases that will be chained to the main SNDS database where possible. In March 2021, the HDH's Scientific Board was asked to give its opinion on the interest of the databases to be included in the first version of the catalogue. The aim was to inform the SNDS Strategic Committee's reflections by providing a scientific opinion on the proposed databases in the absence of an established process, given that the Strategic Committee was only established last July. It therefore determined the scientific value of each database by assessing the originality at national and international levels, the relevance and quality of the database and the strategic research questions that it could potentially address, with or without chaining to health insurance data. The Board evaluated some 30 databases, 18 of which were selected for this first version of the catalogue.

The Unibase programme

The Unibase programme is central to the partnership with Unicancer. In February 2021, the Board was asked to provide an opinion on the scientific value of the databases proposed by the French Cancer Control Centres (CLCC), with a view to their eventual sharing in the HDH catalogue. Based on the information about the types of data available, as provided by Unicancer, the Board assessed the following aspects:

- the potential impact of the proposed topics in terms of mobilising data for academic research and cancer patient care,
- the benefit of chaining the proposed databases to the main SNDS database,
- the relevance of the databases proposed for inclusion in the catalogue and their reusability,
- the relevance of the themes proposed for projects with different aims and methodologies,
- the benefits of data quality and interoperability for other health data actors.

In a second phase, the Board issued its opinion on the themes chosen for the launch of a call for proposals aimed at French Cancer-Control Centres, to enable the provision of support by the HDH for three pilot projects from 2022 onwards.

Involvement of the Health Data Hub at European level

In late autumn 2021, the Board formed a small working group to help the HDH formulate an application for a European Commission call for projects to develop and test an initial version of the European Health Data Space. Drawing on their experience, the group members provided a critical analysis of the choice of European partners, the proposed types of use for research, and the project architecture.

28

MEMBERS OF THE SCIENTIFIC ADVISORY BOARD, 7 OF WHOM ARE INTERNATIONAL

An internationally oriented Scientific Board

The HDH's Scientific Advisory Board is also responsible for providing guidance to the senior management team, based on international sources of inspiration. Seven international members therefore joined the 21 other members in 2021. In general, all of these members, French and international, help the HDH to identify partnerships or new tools and services operating outside France. They contribute to the wider dissemination of knowledge within the global scientific community and to the HDH's integration into the European health data landscape.

Six new data challenges planned for 2022

With its partners, the HDH organises regular unifying events open to the entire international scientific community. The annual Winter School and the Data Challenges are two such examples.

The first data challenge – a "Grand Défi" (Major Challenge) programme entitled "Improving diagnosis with artificial intelligence" – was organised in 2020 with the French Pathology Society, bringing together nearly 550 participants who developed and tested AI algorithms to enable the automatic detection of pre-cancerous and cancerous lesions potentially occurring in the cervix. They based their work on an anonymous dataset of nearly 5,000 histopathology slides (biopsies and conisations) from some twenty French anatomopathology centres, both public and private, which participated on a voluntary basis. The Health Data Hub provided organisational and logistical support for the collection of this data.

At the end of the event, open access to the anonymised data and the algorithms developed is habitually provided to enable the performance of further research.

To encourage such emulation, the HDH has renewed its partnership with the Grand Défi programme and the BPI, and in March 2021, launched a call for expressions of interest aimed at actors in the medical sector. Six data challenges were selected, which are currently supported by the HDH and funded by the BPI.

These initiatives cover various medical fields and will make a significant contribution to the following issues:

1. The DigiLUT Data Challenge, led by Hôpital Foch, to develop an algorithm capable of detecting graft rejection following lung transplants.
2. The Data Challenge, organised by the French Society of Thoracic and Cardiovascular Surgery (SFCTV), to classify lung tumours based on cellular imaging data.
3. The D-IA-GNO-DENT Challenge, led by Strasbourg University Hospitals, to propose algorithms for the automatic diagnosis of rare oral diseases, based on photographs of teeth.
4. The Data Challenge, organised by the French Pathology Society (SFP) and dermatologists, to predict the risk of metastatic melanoma progression.
5. The Allergen Chip Challenge, led by the French Society of Allergology (SFA), to identify allergens and allergen combinations involved in an allergic event.
6. The Data Challenge, led by the French Nuclear Medicine Society (SFMN), to develop an algorithm contributing to the diagnosis of Parkinson syndromes with multiple forms.

Action 3.2 – Structuring and standardising health data to facilitate their use in international collaborations

Towards a standardisation of the SNDS to promote its re-use for research purposes

The SNDS is one of the most comprehensive medico-administrative databases in the world. However, its use requires an expert on the database in order to understand the data schema, its potential, biases and limitations. Using the SNDS to conduct Europe-wide projects remains extremely challenging, due to the use of data schemas and terminologies specific to each database/country. Since 2020, the HDH has been focusing on simplifying the SNDS data schema and adapting it to a common OMOP-CDM data model in order to facilitate its reuse in national and European research projects. Converting the SNDS into OMOP-CDM format means that the database can be condensed from more than 180 tables to fewer than 20 tables. In addition, the SNDS contains more than 70,000 medical terminologies (CCAM, NABM, CIP, UCD, CSARR, LPP, etc.), and work on aligning these terminologies with the standard format is underway with medical interns providing valuable support.

FROM 180 TABLES

TO 14 TABLES IN OMOP FORMAT

40,000 MEDICAL TERMINOLOGIES TO BE ALIGNED

In December 2021, the HDH won its second EHDEN call for projects, which will enable the pursuit of this transformation project and the extension of its scope to all beneficiaries and to the years before 2019.

Action 3.3 – Contributing to the construction of the European Health Data Space based on secondary use of health data

To facilitate access to the different types of data available in its Member States, the European Commission has made the future European Health Data Space (EHDS) a priority of its health policy. This policy of sharing is indeed considered fundamental to the development of research and the improvement of citizens' health. For example, the pooling of data will facilitate the development of more efficient artificial intelligence algorithms to benefit research or industrial performance, and will improve the management of pandemics such as COVID-19, or the study of rare diseases.

Several instruments designed to lay the foundations for the future European Health Data Space have already been put in place by the European Commission since 2019. TEHDaS (Towards a European Health Data Space) is one such example, a reflection programme bringing together more than 26 Member States for which the Health Data Hub is coordinating five affiliated entities: the ANS, INSERM, the Hospices Civils de Lyon (HCL) and the Universities of Aix-Marseille and Toulouse III.

More recently, the Health Data Hub has led a consortium bidding for the European Commission's call for applications to develop a test version of the future European Health Data Space. This consortium brings together national health data platforms, as well as European agencies and research infrastructures. The project consists in rolling out a network of data platforms within the European Union.

Regarding the use of real-world data, the Health Data Hub is also participating in the Scientific Advisory board of the Data Analysis and Real World Interrogation Network (DARWIN) – the data analysis, expertise and service network launched by the European Medicines Agency (EMA), which is aiming to use this data to improve decision-making throughout the life cycle of medicines.

The Health Data Hub is also participating in a consortium led by the start-up Owkin in partnership with INRIA, the Centre Léon Bérard, Nantes University Hospital, Apricity (France), TMF-EV, the Medical University of Berlin, the University of Düsseldorf, BfArM (Germany), Budapest University of Technology and Economics (Hungary), the Medical University of Graz (Austria), the University of Golnik (Slovenia), the Italian National Cancer Foundation, and the Jimenez Diaz Institute (Spain). This consortium has submitted a bid in response to a European call for projects for the implementation of a federated learning system in Europe, taking account of the need for data confidentiality, in which the Health Data Hub would act as a data source node. The result will be announced in March 2022.

First results of the TEHDaS joint action

Launched in February 2021, the joint action "Towards the European Health Data Space" (TEHDaS), mobilising 25 countries, aims to submit recommendations for the future European Health Data Space to the European Commission by 26 July 2022. Led by the HDH, the French contribution is particularly active on the themes of civic engagement (Work Package 8) and governance (Work Package 5). It achieved its first milestones in 2021, through concrete actions.

A major online consultation was launched by the HDH at the end of 2021, together with its Belgian and English partners, in order to measure citizens' perceptions of health data and its use. Available in English, French and Dutch but open to all European citizens, it provides information about data sharing and use in Europe, individual questionnaires and includes a discussion area. In preparation for this e-consultation, an international literature review of nearly seventy articles was conducted, focusing on mechanisms for citizens' engagement in the secondary use of health data in Europe and the re-use of data for purposes other than direct care. User associations, health professionals, public and private health actors and academics were also interviewed.

To consider the future governance of the sharing of cross-border health data, the Health Data Hub and its Finnish, Spanish and Dutch partners have launched an international benchmarking process to compare the procedures for accessing health data in the four countries. The first results have highlighted the barriers to the cross-border sharing of health data and the specific legal requirements for health data.

9% **OF THE EUROPEAN JOINT ACTION TEHDAS BUDGET FOR FRANCE**

60% **FUNDED BY THE EUROPEAN COMMISSION**

The HDH establishes a consortium to propose a test version of the EHDS to the European Commission

Sharing data on a European scale is fundamental to developing research and improving citizens' health. For example, it will enable the development of more powerful artificial intelligence algorithms, to improve the management of pandemics such as COVID-19, and for studying rare diseases.

To this end, the HDH has established a consortium in response to the European Commission's call for applications, published on 25 October 2021. Objective: to establish a test version of the European Health Data Space (EHDS). This consortium brings together some fifteen major European players, including the national health data platforms of several Member States (Finland, Norway, Denmark, Germany, Belgium, Hungary, Croatia), as well as European agencies such as the European Medicines Agency (EMA), the European Centre for Disease Prevention and Control (ECDC) and research infrastructures (BBMRI, Elixir, eBrains, IACS).

Together, they will build a networked technological infrastructure that will interconnect all data-providing partners (data nodes). This network, accessible via a single portal, will enable researchers who are engaged in developing projects to query the data catalogues of all nodes, request access to selected data via a single common form and, if necessary, transfer data to a single location. The principle is to make this process as simple and smooth as possible for project leaders.

The application file submitted on 25 January also includes a series of concrete examples of use which can be used to test the network and demonstrate the benefits of this transnational re-use of health data for research, innovation, health policy, regulation and personalised medicine.

The results of the call for proposals are expected by the summer of 2022 and the winning consortium will start working in September. They will serve as the basis for the future European Health Data Space.

5 **RESEARCH ENTITIES (BBMRI, ELIXIR), ASSOCIATIONS (EUPHA) OR EUROPEAN PARTNER AGENCIES (EMA, ECDC)**

10 **COUNTRIES INVOLVED**

Action 3.4 – Promoting the Health Data Hub internationally

One of the HDH's missions is to contribute to extending France’s international influence in health data uses by promoting the excellence of French institutions. To this end, the Health Data Hub publishes an international newsletter distributed to more than 800 qualified contacts worldwide, and has participated in numerous internationally renowned events. For example, the HDH organised a joint seminar with MIT and the National Academy of Medicine on artificial intelligence in the medical world, and maintained a virtual presence in the French Pavilion at the Dubai World Expo.

The HDH has also signed a partnership contract with its Finnish counterpart, Findata, and initiated exchanges with a view to setting up partnerships with numerous players in Europe and abroad (Canada, Israel, United Kingdom, etc.). The aim of these partnerships is to launch joint reflections on the use of data for research or innovation purposes, and to create impactful data-related projects involving both French and international researchers.

Stand-out events in 2021 also included the organisation of a Winter School dedicated to artificial intelligence in health, co-organised with three Interdisciplinary Institutes for Artificial Intelligence (Prairie, 3IA Côte d'Azur and MIAI Grenoble), which attracted 428 participants, 56% of whom were students. For five days, internationally renowned researchers presented their advances and led practical sessions. The opening lectures were given by Yoshua Bengio (A.M. Turing Award, Scientific Director of Mila and IVADO, Professor at the University of Montreal and Co-Director of the Machine Learning and Biological Program at CIFAR in Canada), I. Glenn Cohen (Professor and Associate Dean of the Petrie-Flom Center for Health Law and Policy, Biotechnology and Bioethics at Harvard University) and Ran Balicer (Professor, Founding Director of the Clalit Research Institute, Director of Innovation at Clalit Health Services and member of the National COVID-19 Expert Committee for the Ministry of Health in Israel).

438

PARTICIPANTS IN THE “AI4HEALTH” 2021 WINTER SCHOOL

The HDH joins the National Academy of Medicine and MIT for a symposium on AI

How to combine Artificial Intelligence and clinical practice? What are the answers to the ethical and regulatory questions raised by AI? The rapprochement between Artificial Intelligence and medicine, facilitated by technological advances, still raises many questions. In an attempt to address these concerns, the HDH, the French National Academy of Medicine and MIT organised a symposium on 5 and 6 May 2021. More than 700 participants, attending the event online, had the opportunity to interact with French and American clinicians and experts, working at the interface between AI and medicine, in order to develop a better understanding of the different approaches and facilitate transatlantic collaboration.

Dedicated to theoretical and technical advances in artificial intelligence applied to medicine, the first day of the symposium focused on the contributions of imaging to guiding medical practice in fields such as the treatment of COVID-19, coronary problems, neurology and oncology. This session gave Nathalie Lassau, Professor of Radiology at Institut Gustave Roussy (IGR), and her colleagues an opportunity to discuss the importance of international consortia in the aggregation of a large number of annotated images. Daniela Rus, Professor of Computer Science at MIT CSAIL, presented the state of the art in robotics for health. The final session was devoted to the contribution of artificial intelligence to the analysis of cell and tissue samples. Cécile Badoual, Professor of Anatomy and Cytology at Université Paris Descartes, presented convincing results on cancer classification and discussed microscopy data and the use of predictive models, particularly in haematology.

The second day, opened by Prof. Bernard Nordlinger, Member of the French National Academy of Medicine, and Cédric Villani, Member of Parliament, was devoted to ethical and regulatory issues related to AI. The difference in approach between the United States and Europe was highlighted, with reference to their respective regulatory frameworks. In a first session devoted to the regulation of data collection and use, Danny Weitzner, Professor of Computer Science at MIT, Jeanne Bossi-Malafosse, a lawyer, and Nicholson Price, Professor of Law at the University of Michigan, showed why health data differs from other types of data and poses major challenges. The adaptation of legislation to these challenges on both sides of the Atlantic was also discussed, as were the shortcomings of the law in certain cases. Laws that may be either overly protective or not sufficiently protective depending on the uses: a source of risk for data protection, or a source of obstacles to research. In the second session, I. Glenn Cohen, Professor of Law at Harvard University, and Claude Kirchner, Director of the French National Pilot Committee for Digital Ethics (CNPEN), addressed the ethical challenges of AI in medicine. During these discussions, I. Glenn Cohen called upon decision-makers to consider the ethical issues raised by the potential biases of artificial intelligence solutions such as adaptive learning.

Finally, Emmanuel Bacry, Scientific Director of the Health Data Hub and Director of Research at the CNRS, and David Sontag, Professor of Computer Science at MIT CSAIL, presented the procedures for accessing health data in France and the United States and the specificities linked to major national programmes. A discussion that highlighted common issues concerning data upgrading, data formats, database interoperability, data openness and relationships with patients, with particular regard to reuse of health data.

704

PARTICIPANTS IN THE "AI AND MEDICINE: PROMISES AND LIMITATIONS" EVENT

PRIORITY 4
ENSURING THE PARTICIPATION OF CIVIL SOCIETY AND PROMOTING HEALTH DATA USE

Before the HDH can achieve its full potential, all stakeholders – and especially the general public – will need to understand its objectives. To this end, the HDH has established a Citizen Relations Department, which is responsible for listening to the needs of civil society and producing content that is relevant to these needs, while also devising partnerships.

Action 4.1 – Listening to civil society and understanding their needs in order to build a relationship based on trust

The Citizen Relations Department carries out actions to understand citizens' perceptions, ascertain their level of knowledge, and gather their needs and expectations concerning health data. To this end, it uses sociological methods, produces literature reviews, leads working groups and proposes consultations with citizens. Several working groups were organised in 2021, including with pharmacy students and the i-Share cohort, to obtain their reactions to proposals for adapted content that meets their expectations. A literature review, contributing to the joint action entitled "Towards a European Health Data Space", was published and an online debate was launched at the European level. All of these initiatives will continue in 2022, in order to inform the activities of the Health Data Hub's Health Data Observatory, and as part of the Open Government commitments.

A major online debate to gather European citizens' views on health data

Citizens' involvement is essential to the process of using health data for research, public health and prevention purposes. To this end, the HDH and its Belgian and UK partners have launched a major online consultation, which will last from the end of December 2021 to April 2022. This ground-breaking consultation is supported by the European Commission, as part of the joint action "Towards a European Health Data Space" (TEHDAS).

Its primary objective is to raise citizens' awareness of how, for what purposes and under what conditions their data is used, and of the role they can potentially play. But beyond that, it is also an opportunity to interact with them, and to foster discussion between citizens themselves. The lessons learned from this debate will inform the proposals submitted to the European Commission in 2022 for the implementation of a common European Health Data Space.

To this end, a dedicated website has been created, containing an educational section to inform citizens, a fun test to arouse their curiosity and gather their perceptions, and a consultation space containing open-ended questions to which they can respond with contributions, comments, "likes" and "shares".

The platform is available in several languages and is open to all European citizens. The HDH and its partners are committed to informing citizens in a neutral and transparent manner, guaranteeing the security and confidentiality of contributions, explaining how they will be used and, of course, taking account of the opinions expressed when formulating recommendations.

In preparation for this consultation, the HDH and its partners conducted a literature review of more than 70 European publications and met with some 60 actors involved with the secondary use of health data.

Action 4.2 – Explaining and disseminating the Health Data Hub's commitments to compliance with procedures, control of uses, management of data, explanation and assistance with the exercise of rights

After the publication of the decree on the French National Health Data System (SNDS), which sets out the new procedures for exercising individual rights, the HDH has updated its website to explain the provisions of the new legislation. The HDH also contributes to the thinking on how to inform the French population as a whole about the existence of the SNDS and their associated rights.

In order to implement individuals' rights, in view of the imminent integration of the main SNDS database, the HDH has launched the design of a rights management tool that may be based on a online form to enable citizens to express their choices once and for all, in a single place (also see: Action 1.2 - Establishing an automated national procedure to facilitate data selection, transmission to the catalogue, and matching, p. 23).

In addition, the HDH has produced documentation for the data controllers of future catalogue databases and the data controllers of projects involving health data. Objective: to help them provide exhaustive, comprehensible information that complies with the GDPR and the SNDS decree. These educational documents aimed at the general public have been designed by lawyers and citizens to enable the people concerned to understand their content.

Finally, through its membership of the network of the French Association of Personal Data Protection Officers (AFCDP), the HDH contributes to a community that engages in discussions on data sharing. It has worked on deliverables with this network, within working groups on the reuse of health data and on the production of "easy-to-read-and-understand" ("Facile à Lire et à Comprendre" – FALC) information.

Action 4.3 – Establishing a policy on communication and on the dissemination of a health data culture

The Health Data Hub contributes to making key issues relating to the reuse of health data understandable and accessible to the general public, and conducts training activities for this purpose. One of these activities sets out to raise citizens' awareness of the National Health Data System (SNDS), with the support of France Assos Santé and L'Assurance Maladie.

The Health Data Hub is also committed to disseminating information about the projects it supports, their aims and their results. Its Citizen Relations Department is committed to the co-construction of this information and the implementation of the "easy-to-read-and-understand" approach. Information actions with health institutions and digital mediators have also been initiated.

Finally, the Citizen Relations Department forges partnerships with civil society and associations to give them an opportunity to become actors in health data. In 2021, for example, it continued to assist France Assos Sante and its member associations (including France Handicap, UNAF, FFAAIR, AFA, EFAPPE, Fibromyalgia France, France Rein, UNAFAM, France Alzheimer, Aides, FFD, FNAR, Ligue Nationale contre le cancer) with the performance of the "Vivre-Covid19" study. The HDH is assisting them with the coordination of the study, the search for additional partnerships with public actors for the exploitation of the data, and the development of a doctrine of transparency and an open-source policy.

Other partnerships have been signed, notably with the FFD, which is the first association to have submitted a project using SNDS data, but also with APODEC, JNA, France Vascularites, TousPartenairesCovid and ANEPF. New partnerships are being discussed.

16
TRAINING COURSES OFFERED ON THE SNDS

"Vivre-Covid 19": an associative study open to research and the scientific community

Launched in May 2020, just 6 weeks after the start of the initial lockdown, Vivre-Covid19 was the first major survey to be conducted by France Assos Santé (FAS), in partnership with the HDH. Each month, it requires more than 5,400 registered participants to describe their anxiety levels and difficulties in accessing care during the health crisis. This study, which ends in May 2022, is open to researchers and to the entire scientific community.

It differs from standard studies because it started during the first lockdown, and because it covers both chronic and disabled patients and a control sample without any pathologies. With more than 3.5 million data items collected by the end of the study, FAS will be able to draw solid conclusions, and has also called on public research partners to exploit the data as widely and precisely as possible.

Its protocol has been posted on the study website, to give all the associative partners and the scientific community a comprehensive insight into its context, objectives, methods, and means, in addition to its security measures, information and analyses. This open-access publication also sets out to demonstrate the scientific and methodological robustness of the study, the relevance of the chosen service providers, and the voluntary sector's capacity to adopt best practices and standards in the interest of patients.

Some of the data will also be freely accessible as anonymised data, to increase the opportunities for new discoveries. However, each entity that uses an anonymised dataset must undertake to use the data for public-health purposes and not for commercial gain.

The HDH is supporting FAS throughout the project, by helping to lead workshops and design and implement the ethical doctrine, for example.

+5,000

PARTICIPANTS IN THE VIVRE COVID-19 STUDY

Action 4.4 – Developing the provision of strictly anonymous data in open-access format

The provision of totally open access to certain data has now made the re-use of health data a widespread practice. To this end, and given the sensitivity of such data, pseudo-data, or completely anonymised real-world data, must be produced. By reflecting some of the properties of real-world data, this pseudo-data facilitates the appropriation of health data and promotes further uses. In order to create this type of large-scale data, the HDH has significantly improved the synthetic data generator, launched in 2020 and freely accessible on Gitlab. A partnership has also been established with INRIA and the REPERES team, to generate synthetic data that is representative of the main SNDS database. These activities will eventually enable the HDH to provide, along with its metadata catalogue, synthetic data to accompany each of the databases, which will be made available on the highly secure platform in pseudonymised form.

The HDH also promotes open access to real-world data when it can be totally anonymised. In 2022, the roll-out of the open-source tool Dataverse will enable users to access open data as part of the supported data challenges.

5,000

DIGITISED UTERINE BIOPSY SLIDES BEING MADE AVAILABLE IN OPEN-DATA FORMAT

3

HEALTH DATA ACCESS REVIEW

To improve the quality of care, health data must be accessible to as many people as possible, while upholding the rights of the individuals concerned. Indeed, these data, even if pseudonymised, remain personal data unless they are totally anonymised. In the latter case, they can be freely used with no formalities required. Otherwise, access to this data is regulated.

ACCESS TO DATA AND OUTLOOK

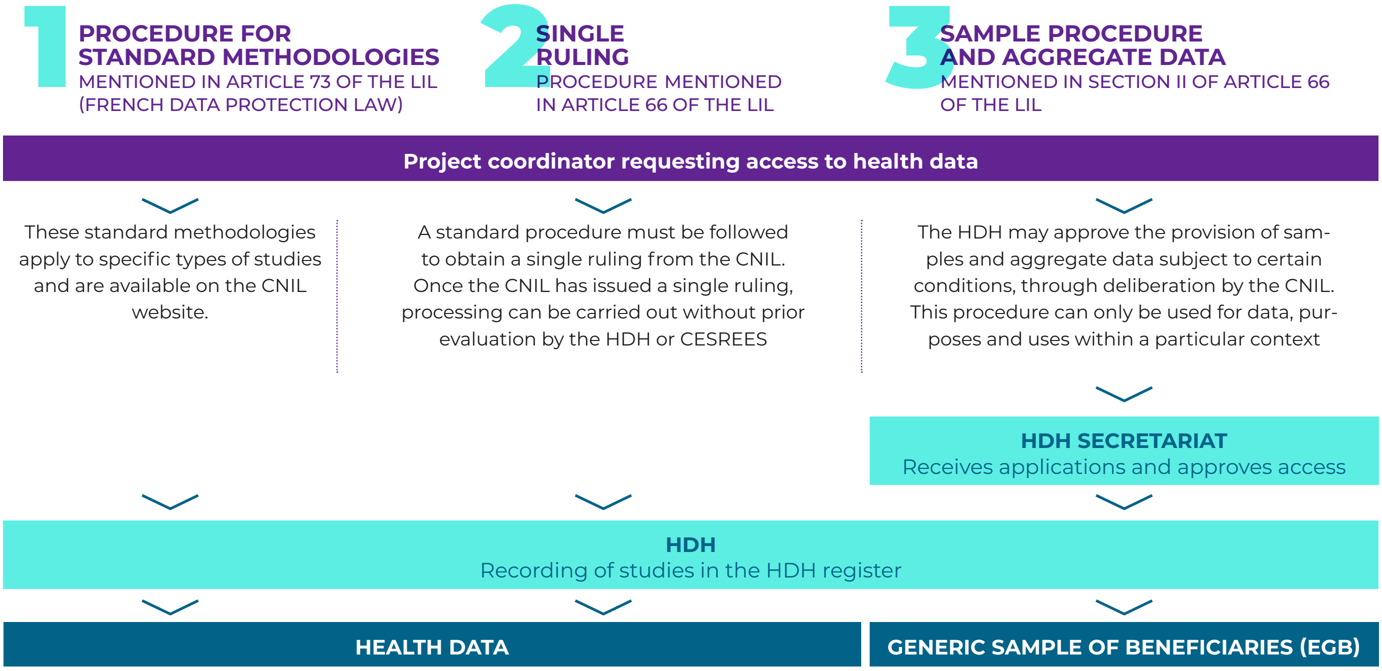
Regulated access

In most cases, authorisation by the French Data Protection Authority (CNIL) is required for processing personal health data. Prior to this authorisation, the procedure provides for the interventions of two institutions:

- The HDH – the single gateway for projects using health data. Nearly **380** projects were submitted to it 2021.
- CESREES (French Scientific and Ethical Committee for Research, Studies and Evaluations in the Health Sector) issues an opinion on the relevance and scientific quality of projects.

In order to facilitate access to the data and allow project leaders to carry out their processing as quickly as possible, the time allocated to each of these institutions is regulated. The HDH therefore has seven days in which to transmit the complete project to CESREES, which then has one month to issue an informed opinion. If this opinion is favourable, the CNIL has two months, possibly renewable, to issue its authorisation. In the event of non-compliance with these time frames, the opinions or decisions will be considered tacitly favourable or granted.

THREE SIMPLIFIED PROCEDURES CORRESPONDING TO USES AND DATA



1 PROCEDURE FOR STANDARD METHODOLOGIES
MENTIONED IN ARTICLE 73 OF THE LIL (FRENCH DATA PROTECTION LAW)

2 SINGLE RULING
PROCEDURE MENTIONED IN ARTICLE 66 OF THE LIL

3 SAMPLE PROCEDURE AND AGGREGATE DATA
MENTIONED IN SECTION II OF ARTICLE 66 OF THE LIL

Project coordinator requesting access to health data

These standard methodologies apply to specific types of studies and are available on the CNIL website.

A standard procedure must be followed to obtain a single ruling from the CNIL. Once the CNIL has issued a single ruling, processing can be carried out without prior evaluation by the HDH or CESREES

The HDH may approve the provision of samples and aggregate data subject to certain conditions, through deliberation by the CNIL. This procedure can only be used for data, purposes and uses within a particular context

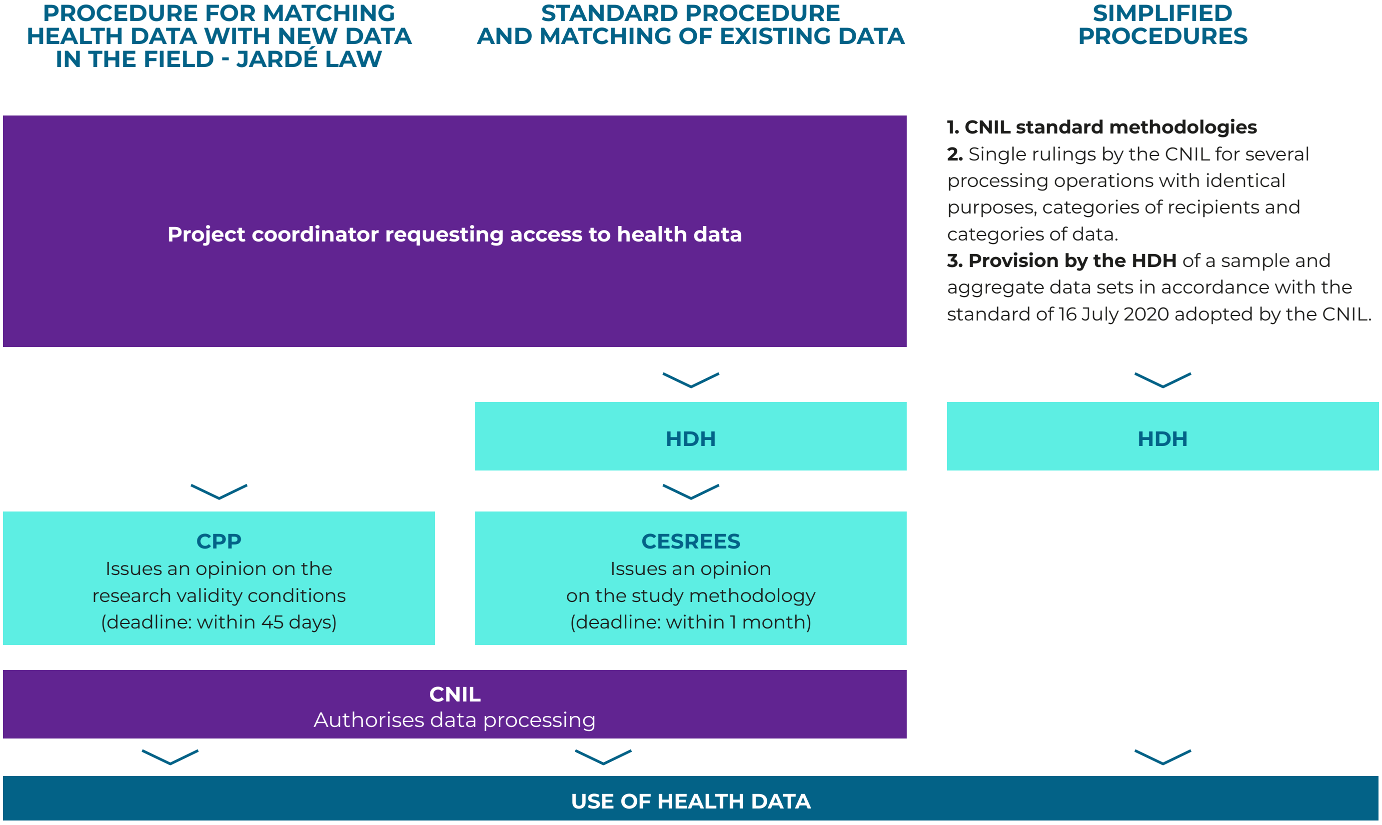
HDH SECRETARIAT
Receives applications and approves access

HDH
Recording of studies in the HDH register

HEALTH DATA

GENERIC SAMPLE OF BENEFICIARIES (EGB)

OUTLINE OF THE HEALTH DATA ACCESS APPLICATION CIRCUIT



PROCEDURE FOR MATCHING HEALTH DATA WITH NEW DATA IN THE FIELD - JARDÉ LAW

STANDARD PROCEDURE AND MATCHING OF EXISTING DATA

SIMPLIFIED PROCEDURES

Project coordinator requesting access to health data

1. CNIL standard methodologies
2. Single rulings by the CNIL for several processing operations with identical purposes, categories of recipients and categories of data.
3. Provision by the HDH of a sample and aggregate data sets in accordance with the standard of 16 July 2020 adopted by the CNIL.

HDH

HDH

CPP
Issues an opinion on the research validity conditions (deadline: within 45 days)

CESREES
Issues an opinion on the study methodology (deadline: within 1 month)

CNIL
Authorises data processing

USE OF HEALTH DATA

"Simplified" access procedures exist, in which CNIL authorisation is not required. This particularly applies to cases in which the intended processing conforms to a standard methodology (méthodologie de référence – MR). To promote data access and use, the HDH is contributing to the roll-out of such procedures in close collaboration with the CNIL. In 2021, for example, it issued an opinion on the standard applicable to health data warehouses. This simplified procedure enables data controllers to set up data warehouses in healthcare institutions without seeking authorisation from the French Data Protection Authority (CNIL),provided that their projects conform to the standard.

Finally, the HDH can also promote easier access to aggregated datasets or samples, under conditions previously authorised by the CNIL.

Key figures on access (2021 figures)

Total number of projects submitted to the Data Access Team in 2021	**380**
Number of projects submitted to the Data Access Team for CNIL authorisation	**366**
number of these projects transferred to CESREES	**279**
number of these projects redirected (standard methodology or Ethics Committee)	**56**
number of incomplete projects	**31**
Number of requests for simplified access to the Generic Sample of Beneficiaries (EGB)	**14**
Number of projects using a standard methodology (MR)	**1,411**
MR004	**1,279**
MR005	**54**
MR006	**78**

A service offering designed for project leaders

Promoting the use of health data requires the documentation of what it contains, the opportunities that it provides, the clarification of the procedures for accessing this data, and the promotion of knowledge sharing. This is why, even before the HDH was created, the Directorate for Research, Studies, Evaluation and Statistics (DREES) developed a collection of operational and educational tools for project leaders. These tools support them at all stages of their project and are improved every year. In 2021, the HDH made the following changes to them:

- The educational kit, which helps project leaders to understand what documents are required for a permit application, has been updated. This kit contains several components, including a guide for completing the scientific protocol, a guide for completing the expression of needs when data from the SNDS core database is required, and a guide relating to the information for individuals required by the GDPR. These components have been designed in conjunction with CESREES and the French National Health Insurance Fund (CNAM).
- A guide presenting the French regulations on access to health data has been created, which improves the actors' understanding by presenting clearly identified criteria. This guide has been translated into English and has been the subject of regular training activities. In addition, a specific guide has been designed to help actors who want to access the Generic Sample of Beneficiaries (Échantillon Généraliste de Bénéficiaires – EGB).
- The range of training courses has been expanded, including training for citizens on the main SNDS database. Aimed at anyone with questions about health data and its use in research projects, this presentation provides an overview of the SNDS and how researchers can use it. Among other issues, it discusses access to health data, its pathways, the concept of open data and the role of the HDH. Each module lasts 20 minutes. This training course is freely available on the HDH website. The training course on SNDS data and its access methods, conducted in collaboration with L'Assurance Maladie, has been reviewed. This course is aimed at all project leaders who are accustomed to using data but may not be familiar with the details of the SNDS, its main databases, the access procedures and the associated regulatory procedures. This training is delivered by video-conference or in face-to-face sessions and lasts for three hours.
- The collaborative documentation providing information for project coordinators on the characteristics and content of the main SNDS database has been enriched.
- 307 questions were asked on our online self-help forum in 2021.
- Seven meet-ups on SNDS-related topics were offered to our 1,639 members.

Health data use and transparency

In 2021, the Public Project Directory, provided for by law, was enriched with 1,800 additional projects. In total, almost 6,000 projects involving personal health data were listed at the end of 2021. An information sheet, specific to each project, presents the expected benefits, the data mobilised, the method, the contact details of the project coordinator and the DPO, and an implementation schedule. A collection of the results of these projects has also been launched, in accordance with the French Public Health Code. To help the actors concerned understand this obligation, an educational guide has been designed and will be published shortly on our website. The results of the projects will also be available for download from the public directory in the coming months.

In 2021, a process of reflection on how to develop the platform for submitting requests for data access was also launched, and will continue in 2022, with a view to improving its responses to the needs of the actors involved in this procedure.

HEALTH
DATA HUB

Easy, unified, transparent and secure access to health data in order to improve the quality of care and patient support.

9 rue Georges Pitard, 75015 Paris

E-mail: hdh@health-data-hub.fr

Publishing: Health Data Hub
Design and production: Parimage
March 2022