

The French Health Data Hub and the German Medical Informatics Initiatives Two National Projects to Promote Data Sharing in Healthcare

Marc Cuggia, Stéphanie Combes

▶ **To cite this version:**

Marc Cuggia, Stéphanie Combes. The French Health Data Hub and the German Medical Informatics Initiatives Two National Projects to Promote Data Sharing in Healthcare. IMIA Yearbook of Medical Informatics, 2019, 28 (1), pp.195-202. 10.1055/s-0039-1677917 . hal-02280247

HAL Id: hal-02280247

https://univ-rennes.hal.science/hal-02280247

Submitted on 8 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Published online: 2019-08-16

© 2019 IMIA and Georg Thieme Verlag KG

The French Health Data Hub and the German Medical Informatics Initiatives: Two National Projects to Promote Data Sharing in Healthcare

Marc Cuggia[1], Stéphanie Combes[2]
[1] INSERM, UMR 1099, Rennes, France and Université de Rennes 1, LTSI, Rennes, France
[2] Lab Santé, Sous-direction de l'observation de la santé et l'assurance maladie, DREES, France

Summary

Objective: The diversity and volume of health data have been rapidly increasing in recent years. While such big data hold significant promise for accelerating discovery, data use entails many challenges including the need for adequate computational infrastructure and secure processes for data sharing and access. In Europe, two nationwide projects have been launched recently to support these objectives. This paper compares the French Health Data Hub initiative (HDH) to the German Medical Informatics Initiatives (MII).

Method: We analysed the projects according to the following criteria: (i) Global approach and ambitions, (ii) Use cases, (iii) Governance and organization, (iv) Technical aspects and interoperability, and (v) Data privacy access/data governance.

Results: The French and German projects share the same objectives but are different in terms of methodologies. The HDH project is based on a top-down approach and focuses on a shared computational infrastructure, providing tools and services to speed projects between data producers and data users. The MII project is based on a bottom-up approach and relies on four consortia including academic hospitals, universities, and private partners.

Conclusion: Both projects could benefit from each other. A Franco-German cooperation, extended to other countries of the European Union with similar initiatives, should allow sharing and strengthening efforts in a strategic area where competition from other countries has increased.

Keywords

Big data, interoperability, data sharing

Yearb Med Inform 2019:195-202
http://dx.doi.org/10.1055/s-0039-1677917

1 Introduction

Everywhere in the world, ambitious projects for exploiting the tremendous potential of health big data (HBD) by different actors (researchers, physicians, healthcare institutions, citizens, start-ups, R & D laboratories of digital technologies, and patient action groups) are constantly arising. HBD holds the promise of supporting a wide range of medical and healthcare functions, e.g., clinical decision-making, disease surveillance, and population health management. This explains the incentive policy to derive knowledge from HBD by public authorities and scientific communities [1].

During recent years, developed countries have launched ambitious initiatives to promote open data and data sharing with the goal to increase cooperation among institutions. Examples of such networks are the U.S. Observational Health Data Sciences and Informatics (OHDSI) collaboration [2, 3], the PCORnet Clinical Data Research Networks (CDRNs) [4], Patient Powered Research Networks (PPRNs) [5], and the eMerge network [6, 7]. Similarly, China has launched a national program to enhance HBD utility to enter the next level of knowledge discovery [8].

Europe has also launched several projects at both national and EU-wide levels for sharing metadata or aggregated data around specific topics, such as rare diseases and biobanks [9-11], to sustain clinical research [12], and for creating a European infrastructure to manage and safeguard the increasing volume of data generated by publicly funded research [13]. Conversely, only a few breakthrough initiatives have focused on HBD sharing for innovation. The two newest national scale investment programs are: in France, the nationwide data platform called "Health Data Hub" (HDH) [14] set up to facilitate HBD sharing and exploitation with the aim of fostering the development of AI projects in the health field; in Germany, the national Medical Informatics Initiative (MII) launched by the German Federal Ministry of Education and Research, which plans to develop infrastructure for integrating clinical data from patient care and medical research [15].

In this article, we present the stakes, ambitions, and orientations of the HDH and MII projects, and discuss their similarities and specificities from the perspective of a potential for European cooperation in this domain.

2 Materials and Methods

The two authors agreed to define from the literature and their own experiences the main criteria to compare the HDH and MII approaches. Moreover, the two authors participated in the HDH Committee in charge of establishing the HDH roadmap from April to October 2018. Specifically, from May to September 2018, the Committee was in charge of the organization of 114 meetings and 5 workshops with most of the French health information technology (HIT) representatives. This Committee produced a ministerial report (on which the authors rely to describe the French HDH system), and helped to launch the project at a national level at the end of 2018. One author (MC) also coordinates a national working group

on HBD that involves most of the French academic hospitals. He also manages the Western Clinical Data Center, one of the first Clinical Data Network in France [16].

Concerning the German initiatives, the authors analyzed the five articles [17-21] published in the special issue of *Methods of information in Medicine* published in 2018. One of these articles provides a global view of the German MII project [17], whereas the other four describe the four consortia selected by the German Federal Ministry of Education and Research to build the system.

The authors analyzed and compared the French and German initiatives on the basis of criteria that according to literature data allow and promote health data sharing among stakeholders for data reuse: (i) Global approach and ambitions, (ii) Use cases, (iii) Governance and organization, (iv) Technical aspects and interoperability, and (v) Data privacy access/data governance.

3 Results

3.1 Global Approach and Ambitions

The German MII started in 2015 with national funding and a call for proposals published by the German Federal Ministry of Education and Research. The global aim was to develop innovative IT solutions to improve patient care and research and to harness the opportunities created by digitalization in medicine. The call stressed the need to support and improve data exchange and use across the boundaries of individual institutions and sites, particularly between healthcare and clinical/biomedical research actors. Besides these goals, the German strategy was to promote training in rare and much needed skills and profiles in the ecosystem (e.g., data scientists, data stewards, data privacy officers). The MII call was also designed to promote and strengthen education, training, and professional development in medical informatics at German universities. In the German healthcare system, policy decision-making is shared between the national (federation) and state (Länder actors, with much power delegated to self-governing bodies) levels. In this federal organization, similar initiatives had already been started in single states in recent years. Therefore, the MII call wanted to provide a fresh impetus to develop the e-health strategy across the individual German states, and to federate the existing projects with a bottom-up approach. The German initiative was organized in three steps. First, during the conceptual phase, seven consortia were established that, following an initial external assessment, developed competitive ideas and solutions for setting up Data Integration Centers (DICs) and innovative use cases. At the end of this nine-month phase (August 1, 2016 to April 30, 2017), these consortia were evaluated by international panels and four consortia were selected as the most attractive solutions that should be implemented. The second phase (2018-2021) is focused on the development and networking of these four consortia (DIFUTURE, HiGHmed, MIRACUM, and SMITH) with funding totaling 120 million euros. Germany will give 30 million euros to enable the integration of the other sites into these four consortia. MII consolidation and further development will be carried out from 2022 to 2025 (third phase). The MII project focuses on German university hospitals and medical centers which treat a total of 1.8 million patients annually (10% of all inpatient cases in the country).

The French Health Data Hub ("*Plateforme nationale des données de santé*", HDH) is born from the political willingness to promote artificial intelligence (AI) in health, identified by the French government [16] as a priority domain where AI technologies could provide a strategic advantage for the nation and for Europe [22]. In 2018, President Macron announced two major decisions [23]: the construction of four high-level AI research institutes (*Institut 3IA*), and the creation of the HDH associated with the extension of the National Health Data System ("*Système National des Données de Santé"*, SNDS). SNDS was created in 2015 to merge the national anonymous claim database (SNIIR-AM) with relevant data sources, such as the national diagnosis-related group database (PMSI) and the national death registry (CepiDC) [24]. SNDS covers 98.8% of the French population, more than 66 million persons, making it possibly the world's largest continuous homogeneous claim database [25]. SNDS will be enriched by adding new data, especially clinical data from hospital and primary care electronic health records (EHRs), and will provide tools and services to the different stakeholders (i.e., health data producers, users, and citizens) through a national technological hub (i.e., HDH). HDH's role is to facilitate data sharing and exploitation in high-level security conditions. Thus, from May to October 2018, a specific Committee worked on the definition of a roadmap for this platform and the health minister entrusted the creation of HDH to the director of the Directorate of Research, Studies, Evaluation and Statistics of the Ministry of Solidarity and Health (DREES) [16]. The French initiative is based on a top-down approach. To avoid the dispersion of funds and technological resources across the French territory, the choice that has been made is to offer to data producers and users a critical-size data sharing infrastructure, with cutting-edge technology run by a highly skilled staff (engineers, developers, data specialists, and other medical, legal, or cybersecurity experts) often difficult to hire by local institutions, such as academic hospitals. In this context, data producers and users are considered broadly. They can be academic or private institutions, healthcare industries or start-ups, big pharma, and safety agencies. Sometimes, the same entity can play both roles, for example, academic hospitals and public health agencies. In 2019, France allocated about 10 million euros to HDH conceptual phase. Pilot data sharing projects will be carried out mainly with academic partners. From 2020 to 2022, 76 million euros have been earmarked to complete HDH and to support projects with the whole ecosystem.

3.2 Use Cases

The German initiative gave much freedom to the consortia for the choice of use cases. The four selected consortia have proposed a wide range of domains and application areas that are summarized in Table 1. The aim of these use cases is mainly to allow the consortia to focus on the critical requirements to achieve health data sharing, particularly implementing an electronic consent declaration, implementing a trusted third party for identity

management, define the rules of data use and access, data protection, semantic interoperability, data sharing modalities, audit criteria, patient involvement and empowerment, medical informatics research and education development, roll-out, and further integration.

Regarding the French HDH, during the meetings with HIT players, more than a hundred use cases have been identified in four main domains: medical research, patients' information, decision support systems for health professionals, and public health informatics for health system monitoring. At the beginning of 2019, two groups of pilot projects were selected. The first one included only public or academics partners, while the second one was open also to start-ups and private-sector partners through a call of interest. Finally, 20 projects were selected that covered a wide range of domains. Table 2 provides a global view of these projects. Most of these projects link data coming from the current SNDS with specific registry data. Some projects (i.e., HugoShare and EpifracTAL) aim at crossing real-life data coming from clinical data warehouses with SNDS data. Like the German initiative, all these projects will contribute to develop and complete HDH and its services during the period 2019-2022.

Table 1 Use cases covered by the four MII consortia

HiGHmed	Oncology: Clinical use of genome sequencing Cardiology: Continuous longitudinal monitoring of physical activity Infection control: Cross-site analysis of patient movement data
MIRACUM	Patient recruitment in clinical trials Clinical molecular knowledge tools Supporting molecular tumor boards
DIFUTURE	Infrastructure for personalized optimal treatment of Multiple Sclerosis Infrastructure supporting research on early diagnosis of the different subtypes of Parkinson's disease Application on indolent lymphoma and checkpoint inhibitor therapies in skin and lung cancer
SMITH	Phenotype pipeline to systematically annotate and analyze patient-related phenotypes Algorithmic surveillance of ICU Patients Clinical decision support systems to improve clinical outcomes on patients with bloodstream infections

Table 2 Pilot projects of the French HDH initiative

Pilot projects lead by public-sector partners and academic actors		
SMPP	Oncology/urology	Association of prostatic cancer and metabolic syndrome
ENEAS	Public health/Economy	Access to resources for people with disabilities
BACTERIEMIES	Infectious diseases	Antibiotic resistance during bacteremia
COMBICANCER	Pharmacology	Drug interactions during breast cancer treatment
NHANCE	Clinical decision support systems	Automatic detection of digestive organ lesions
FREISA	Care assessment	Care effectiveness following low-trauma fracture
MULTIMO	Oncology/Ear-nose-throat	Automatic detection of HPV infection in epidermoid tumors
HUGOSHARE	Medication monitoring	Adverse effect detection and medication reconciliation
BNDMR	Epidemiology	Enrichment of the national rare disease registry using SNDS
Pilot projects lead by private-sector partners and start-ups		
DEEPSARC	Public Health/Oncology	Enrichment of the national sarcoma registry using SNDS information
PARCOURS IDM	Public Health/Cardiology	Care pathways of patients with heart failure
HYDRO	Cardiology	Prevention of congestive heart failure crisis
NS-PARK	Public Health/Neurology	Care pathway prediction for Parkinsonian patients
ARAC	Health economics	Calculation of the amount to be paid after reimbursement
ORDEI	Pharmaco-epidemiology	Drug adverse effect detection
OSCOUR	Emergency medicine	Study of stroke events in an emergency unit
DEEP.PISTE	Oncology	AI for breast cancer screening
REXETRIS	Pharmaco-epidemiology	Effects of long-term exposure to immunosuppressive drugs
PIMPON	Drug decision support	Drug interaction alert system

3.3 Governance and Organizational Aspects

MII structuring and funding efforts are focused toward university hospitals and medical centers. MII consider that these entities are the significant starting points, relevant to patient care, and are tightly integrated with medical research on an internationally competitive level. The national steering committee (NSC) leads an action plan to provide assistance and coordination to the four scientific consortia. A coordination office operated by prestigious institutes (Technology, Methods and Infrastructure for Networked Medical Research, German Association of Medical Faculties, and German Association of Academic Medical Centers) supports the NSC action plan. Locally, each consortium must establish DICs and innovative use cases. NSC encourages collaboration among sites and consortia to promote their coordination and interoperability across Germany. NSC is also in charge of coordinating working groups on topics of the national roadmap (e.g., interoperability, standardized data access criteria, data security mechanisms). All university hospital sites in Germany can take part in these working groups that are open to external experts and institutions e.g. the Health Level 7 (HL7) standards development organization, and Integrating the Healthcare Enterprise (IHE). The four consortia have a similar governance and structure, most including a supervisory and an executive board, a technical board, and a use and access committee. In addition, consortia have created local working groups on

key issues, such as ethics and privacy policy, because the regulatory framework and cultural habits differ between the German states. DICs are the pillars of this organization. They are a new structure within academic medical centers, and could be described as local and central hubs to offer expertise in health data science and medical informatics, and provide secure and reliable access to high quality data integrated from healthcare and research sources. Each DIC is dedicated to a specific area, for example, OMICS in the HiGHmed consortium.

In France, the HDH is currently putting in place a governance structure. The Committee had noted that access to each health database was governed by heterogeneous, sometimes complex rules. Indeed, requests for access and matching different sources could take several years to complete. The different actors consulted during the Committee work period had expressed the need for a single access point to ensure simplified, effective, and rapid access to data. The Committee recognized that the harmonization of data access from the existing sources would be highly desirable, but it would require a huge effort to avoid a proprietary reflex or discretionary access to data by the producer. Therefore, they suggested that data producers and users should share data for innovative projects on a secure third-party platform, and that each step of the project should be given adequate support. A second lever was to extend the SNDS by law to all health data encompassed by the welfare system. The goal is to give the same legal status to all health data, independently from their origin. The French HDH plans to leverage and encourage the existing local initiatives by i) creating local or territorial hubs that provide local medical expertise and healthcare services like at the national level; and by ii) encouraging and funding the local efforts (e.g., by public or private hospitals) to provide high-quality data and expertise. From a legal point of view, HDH will be a Public Interest Group by the end of 2019. Meanwhile, the project is currently led by a national steering committee, a scientific board, and a technical board. Like with the German initiative, working groups, open to voluntary contributions, are created to address the different challenges.

3.4 Technical Aspects and Interoperability

Reusing healthcare data for research and sharing data across institutional boundaries immediately raise questions of interoperability and data quality. Regarding the infrastructure, despite some similarities, the strategy seems different from one consortium to the other. HiGHmed should rely on the openEHR platform and components. MIRACUM pursues the strategy to package and extend some existing OpenSource components and software to build its infrastructure, called Miracolix (Medical Informatics ReusAble eCo-system of Open source Linkable and Interoperable) [30]. DIFUTURE will be based on clinical data warehouses (I2B2 [31] for clinical data and Transmart [32] for OMICS data) that are also used by MIRACUM. Conversely, SMITH and HiGHmed designed their platform using a distributed approach based on API and services to connect data sources of legacy information systems to the platform. In this same logic, SMITH offers two high-security level services, one for data sharing, and another one (the SMITH Market Place) dedicated to contracting [33] and granting access to data for internal and external data users. To ensure care coordination and interoperability, the choice made by the German initiative is to define a minimum core dataset to be transferred from the point of care to the DICs for each individual patient case, all in a standardized pseudonymized form for future use in research. The strategy is then to progressively broaden the core dataset according to mutually agreed upon criteria: relevance to research and patient care, relevance to the use cases, availability and accessibility in the MII sites, extent of data structuring, and availability of reference terminologies. The first version of this national dataset is limited to diagnosis and procedures (using the ICD-10 and the German OPS coding systems), but should be extended to laboratory data (by standardization of a subset of laboratory tests coded with LOINC [26]) and drug prescriptions (coded with a national medication terminology). However, comparison of the four consortia suggests that the interoperability strategy and the local implementation are different or at least not fully settled yet. For instance, HiGHmed proposes to extensively use openEHR [27] as the interoperability framework, while SMITH and DIFUTURE are building their strategy mainly on IHE integration profiles and HL7/FHIR [28]. MIRACUM focuses on OMOP as the starting point framework, but will use HL7/FHIR and IHE for clinical data sources. Only the articles on DIFUTURE and SMITH cite DICOM for imaging. HiGHmed and SMITH rely on the CTS-2 terminology services [29]. Concerning the narrative clinical texts, three consortia use NLP methods for information extraction.

In France, the HDH strategy concerning the infrastructrure is to leverage existing initiatives and to rely on the emerging local platforms to build the national infrastructure. Clinical data warehouses are thriving in some academic hospitals. For instance, the University Paris hospitals (AP-HP) are developing the AP-HP Clinical Data Repository, a big data platform that integrates all the clinical data sources of the hospital group and are initializing Data Quality programs to assess the characteristics and the quality of the data [44, 45]. In the same way, the French Cancer Center Federation has been extending a data warehouse network in France [46]. University hospitals in the West of France have built one of the first clinical data research network in Europe [47], which is very similar to a consortia of the German project. After a first phase dedicated to building the core platform to exploit the national data collections and provide a first group of services nationwide, the French HDH plans to decentralize the platform by building local hubs. The aim is to sustain the local efforts in collecting and exploiting HBD, near clinician's expertise while connecting these platforms to the national network. The vision is not to build a centralized infrastructure collecting all health data in a single point but to create a national hub to connect local and national data and infrastructures. HDH will provide IT resources based on cloud computing with a high security level, and will share tools and services that would probably have been too expensive for local hospitals. Among these services, a central catalog will describe all the available data collections and will provide access to data samples for training or testing. Other crit-

ical IT components and services will be developed, such as a national patient portal to manage patient information and consent. In this persepctive, the French initiative encourages data producers to improve data quality by adopting and implementing interoperability standards. Hence, it has been proven that interoperability is a key factor to reach meaningful use of health data, and to improve data reuse [34, 38]. Adoption of common data models and semantics based on formal, accessible, and broadly adopted standards participates to the FAIR principles and improves the automation of health data at a large scale [33]. This strategy has a direct impact on data standardization, with benefits for data reuse and also for patient care. In France, the lack of interoperability in health information systems has been a recurrent and multifactorial problem for many years [39]. Data sources are fragmented in a myriad of software components that are not fully interoperable. Moreover, some vendors do not make their systems fully open, and hospitals struggle to extract data from these systems or to impose the use of standards on their vendors. However, this situation is changing in the right direction. In recent years the national agency for shared health information system (ASIP-Santé) has specified a national interoperability framework that vendors are expected to implement, and health facilities and health professionals are supposed to implement [40]. ASIP-Santé has proposed a national terminology service (i.e. a set of tools and resources) to facilitate the adoption of reference terminologies. During the work of the HDH Committee, the main stakeholders (vendors, associations, such as InteropSanté that promotes IHE and HL7/FHIR adoption) were consulted to define the best incentive measures for the adoption of interoperability standards. Among these measures, the Committee recommended i) the adoption of key reference terminologies and especially SNOMED CT (which is increasingly implemented over the world [41]) to standardize value sets for clinical domains; ii) to make mandatory that medical software are open to allow data availability to customers; and iii) to put the focus of standardization efforts in priority on data sources that can be shared. For instance, a previous national project for implementing CPOE systems in French hospitals allowed the standardization of digitized drug prescriptions nationwide. With the aim of extending the national system for health data, this data source could be easily collected and shared. On the other hand, few hospitals have normalized laboratory data with LOINC, but this is now changing thanks to the work of other national projects with the aim of sharing healthcare data, such as the national project for a shared patient record (DMP) [42] or the project of grouping hospitals in territories (GHT) [43].

3.5 Data Privacy and Data Access Governance

Data privacy and regulatory compliance are very important for building and maintaining patient trust.

In Germany, data privacy oversight agencies as well as the applicable legislation and interpretation regarding whether and to what extent health data can be used in research, and whether informed patient consent is required, vary from state to state. Consequently, the MII consent working group defined a common document and procedures for patient information and consent that respect the local data protection policies, but also allow data sharing and use by multiple parties nationwide. The Federal and the German state agencies validated the resources, documents, and procedures that were jointly developed with national IT agencies. Regarding governance, a common set of rules defines data access procedures across sites and consortia. For instance, use and access committees and transfer offices are mandatory at all DICs with transparent rules on data use requests, and with secured procedures for data transfer. At the national level, an independent central coordination and registration office that receives and manages central data requests ensures transparency on the purpose and achievements of data use toward the public and stakeholders. From a technical point of view, the DICs comply with common security measures. For example, patient identifying information is separated from the clinical and research "payload" data, stored separately, and managed by specific personnel in charge of matching several data sources for the same patient. The four consortia propose different solutions to address data protection and governance issues. For instance, HiGHmed has defined a process at the early stage of a request to perform feasibility studies (i.e., aggregate counts grouped by sites) that does not need patient-specific consent. DIFUTURE has defined the role of a data protection specialist, a DIC member who reviews each request and analyzes the risks of data sharing. MIRACUM has proposed to use an open-source tool for managing consents and data use agreements that was previously developed for consent management of biological samples [48].

In France, the situation is different because data protection is regulated by a national law. The national agency for data protection (CNIL, [49]) is an independent administrative regulatory body that ensures the application of the data privacy law for the collection, storage, and use of personal data. However, the possibility of large-scale data sharing offered by HDH will lead to the development of new and harmonized procedures of data access and to strengthen SNDS security and data protection. The action plan for this issue encompasses a revision of the law, and the choice of state-of-the-art technologies to address the security aspect of the platform. Currently, health data contained in the SNDS can be accessed without the need of patient consent, upon the agreement of a dedicated national committee (CEREES, [50]) and the CNIL authorization. Only public interest studies are allowed. However, the adoption of the European General Data Protection Regulation to enhance patient and citizen information and consent will be imperative. The perspective of crossing multiple data sources at large scale will increase the risk of identification. HDH plans to deploy a portal to provide information to citizens and patients, and to collect their consent. Regarding the local initiatives, there are data access procedures that rely on local committees to assess and authorize each data request on the basis of scientific, ethical, and deontological criteria. Patient consent is currently based on the opt-out approach. Harmonization of access to local data through a national data governance system is one of the great HDH challenges. It is about finding

a point of balance between a strong centralization, which would generate a feeling of loss of control or dispossession for citizens and data producers, and the fragmentation of data access policies, which can lead to discretionary rules and hinder data sharing. The HDH Committee suggested building a strategy based on the subsidiarity principle whereby a central authority should have a subsidiary function, thus performing only the tasks that cannot be done locally [51]. The key idea is to allow the right decision entity to give access to data according to the extent of data sharing. In this model, each institution (e.g., hospital) keeps the control of its own data for projects that require only this data source. If a project needs data coming from different sites, then an inter-site committee will review the project application. A national committee would review the project application in case the project requires collecting data nationwide (e.g., pandemics).

4 Discussion

In this paper, we compared two recent national strategies to promote and strengthen health big data sharing for medical research and innovation. This work has some limitations. We considered only a few criteria and did not discuss some other important aspects, such as the economic model, the relationships with the industrial stakeholders, and the underlying political aspects of these strategies (e.g., national sovereignty issues, position relative to the big tech companies, etc.). We focused on the French and German projects, but other European countries also have started important projects in this domain. For instance, Switzerland has recently invested 1 million Swiss francs per hospital to promote the creation of health data centers [48], similar to the German DICs or the Clinical Data Centers in France. In 2018, the UK government published a national strategy to support AI development and health data sharing [53]. This included a code of conduct for data-driven healthcare technology [54], data collections, and a public campaign called "Your Data Matters to the NHS". The Belgian government approved a pilot project called the Institute of Analytics for Health (INAH) in Wallonia [55]. Its aim is to facilitate the development of innovative therapeutic solutions and to strengthen medical prevention and early identification of pathologies. INAH wants to become the single portal for health data processing and the guarantor of compliance with the ethical constraints of data protection.

We are not planning to list all current initiatives, but only to define a wish list to improve HBD sharing and use on the basis of the strengths and weaknesses of the German and French projects:

a) Combining top-down and bottom-up approaches to build national projects: Data driven life science requires to shift from a competitive paradigm towards a collaborative paradigm. Comparing the two projects, we see that the territorial organization of the country influences the project philosophy and structuration. The French top-down approach is potentially a success factor to mutualize efforts and investments in building the technical infrastructure, to commit stakeholders at adopting interoperability standards and harmonized data sharing procedures. It is also a way to keep a national vision and to define an ambitious roadmap. On the other hand, the German approach stresses the importance of territories and especially the crucial role of local university hospitals in terms of innovation, relationship with patients, and medical expertise, including medical informatics, while having a national coordination. The German initiative gives more freedom to the consortia that could hamper an effective collaboration across the country if technical and organizational choices are too different. The role and the authority of the German National Steering Committee will be therefore crucial for driving and imposing the national roadmap. In contrast, supporting and speeding-up the development of DICs especially in university hospitals, as it is done in Germany, should be a factor of success for the French project. Providing a German federal data sharing infrastructure similar to the HDH might boost the possibilities of data driven and AI projects on a larger, and ideally at a cross-border, scale.
b) Addressing the interoperability as a major European stake: For all the reasons we cited, interoperability and data quality will be essential for extracting reliable knowledge and for the effective use of HBD. MII and HDH initiatives cannot on their own address this critical aspect. For instance, the lack of native language reference terminologies and of national or European services, the dispersion and the insufficient coordination of e-health strategies might maintain the interoperability level of our health information systems at the current low level. A research effort in this domain is also needed as well as the involvement of all institutional and industrial players. Such challenges are currently tackled by initiatives such as the i~HD institute which is a European not-for-profit body. I~HD addresses obstacles and opportunities to using health data by collating, developing, and promoting best practices in information governance and in semantic interoperability [56].
c) Promoting NLP research and development: If we exclude registries or research databases, it is striking to note that structured clinical data in both French and German projects are limited to diagnosis, procedures, and drugs. In contrast, 80% of EHR contents are in free text, and contain a large part of the critically useful information. Therefore, NLP methods that are based on AI are indispensable to leverage this data source, and NLP research must be supported at national and European scales [57].
d) Involving the citizen as a key player and establishing a relationship of trust: HBD and AI technologies hold the promise of innovation and progress for the patient's benefit. However, they also are a potential "Pandora's box" in case of misuse. The cornerstone of data sharing and reuse is trust. Therefore, implementing a trustworthy process for handling citizens' and patients' health data is essential. A system for data reuse should prove its trustworthiness by fulfilling the responsibility of dealing with data within the limits of a social contract regulated by policies between citizens and the organizations that manage the system. Designing a governance system that includes the

subsidiary principle and the active participation of patients and citizens should strengthen the trust toward national projects. Consent management, information, and education actions are key factors to commit citizens. In this regard, the CLARET project of Sherbrooke University in Canada is an exemplary initiative [58, 59]. It proposes alternative methods of information and consent that would be better adapted to the various activities of a learning health system. However, the implementation of such measures will be time-consuming, costly, and complex.

e) Promoting training courses in health data sciences in medical schools: The German initiative promotes training in data science throughout the curriculum of health students. This idea should be adopted by the French national plan by significantly increasing the number of university teachers in medical informatics and by developing easily shareable e-learning resources.

Beyond these factors, any digital transformation needs a strong political leadership and the coordination of involved national institutions. A report [60], recently published, comparing health digital strategies at an international level, shows disparities across countries in their capacity to carry out the digital transformation required to share and leverage HBD. This report concludes that countries which succeed such transformation had adopted a health benefit-oriented and pragmatic strategy. They payed a particular attention to the acceptance among patients, doctors, and other professionals by systematically involving representatives in co-designing strategies and applications. The French and German initiatives seem to comply with these principles.

5 Conclusions

In this paper, we analysed and compared two major breakthrough projects recently launched by the French and German governments to promote health data sharing and support artificial intelligence development in healthcare. Both projects share the same objectives, but differ in terms of strategies. The previous decade has seen major national investments in many European Union and other countries to promote the development of national eHealth infrastructures, which disappointed results as compared to the promises made and expectations. It could be argued that there is no strong evidence of the best way to achieve these current infrastructures at a national this scale. A cooperation between the top down and the bottom up approaches would be helpful to share learning and help discovering emerging best practices.

In this spirit, a Franco-German cooperation, extended to other countries of the European Union with similar initiatives would be valuable to allow sharing efforts and experiences in a strategic area where international competition will become stronger and stronger.

Acknowledgement

The authors would like to thank Christel Daniel and Dipak Kalra for carefully reading their manuscript.

References

1. Bourne PE, Bonazzi V, Dunn M, Green ED, Guyer M, Komatsoulis G, et al. The NIH Big Data to Knowledge (BD2K) initiative. J Am Med Inform Assoc J 2015;22(6):1114.
2. OHDSI – Observational Health Data Sciences and Informatics. [Online]. Available: https://ohdsi.org/. [Accessed: 05-May-2019].
3. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform 2015;216:574–8.
4. Collins FS, Hudson KL, Briggs JP, Lauer MS. PCORnet: turning a dream into reality. J Am Med Inform Assoc 2014;21(4):576–7.
5. Workman TA. Engaging Patients in Information Sharing and Data Collection: The Role of Patient-Powered Registries and Research Networks. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013.
6. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al; eMERGE Network. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med 2013;15(10):761–71.
7. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB et al; eMERGE Team. The eMERGE Network: A consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med Genomics 2011;4:13.
8. Zhang L, Wang H, Li Q, Zhao M-H, Zhan Q-M. Big data and medical research in China. BMJ 2018;360:j5910.
9. Larsson A. The Need for Research Infrastructures: A Narrative Review of Large-Scale Research Infrastructures in Biobanking. Biopreserv Biobank 2017 Aug;15(4):375-83.
10. Bouzille G, Jouhet V, Turlin B, Clement B, Desille M, Riou C, et al. Integrating Biobank Data into a Clinical Data Research Network: The IBCB Project, Stud Health Technol Inform 2018;247:16–20.
11. Merino-Martinez R, Norlin L, van Enckevort D, Anton G, Schuffenhauer S, Silander K, et al. Toward Global Biobank Integration by Implementation of the Minimum Information About BIobank Data Sharing (MIABIS 2.0 Core). Biopreserv Biobank 2016 Aug;14(4):298-306.
12. De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, et al. Using electronic health records for clinical research: the case of the EHR4CR project. J Biomed Inform 2015 Feb;53:162-73.
13. Blomberg N, Oliveira A, Mons B, Persson B, Jonassen I. The ELIXIR channel in F1000Research. F1000Research 2015;4.
14. Healthdatahub. [Online]. Available: https://www.health-data-hub.fr/. [Accessed: 05-May-2019].
15. About the initiative | Medical Informatics Initiative. [Online]. Available: https://www.medizininformatik-initiative.de/en/about-initiative. [Accessed: 05-May-2019].
16. Health Data Hub - Ministère des Solidarités et de la Santé. [Online]. Available: https://drees.solidarites-sante.gouv.fr/etudes-et-statistiques/acces-aux-donnees-de-sante/article/health-data-hub. [Accessed: 05-May-2019].
17. Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med 2018;57(S 01):e50–e56.
18. Prokosch HU, Acker T, Bernarding J, Binder H, Boeker M, Boerries M, et al. MIRACUM: Medical Informatics in Research and Care in University Medicine. Methods Inf Med 2018 Jul;57(S 01):e82-e91.
19. Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O, et al. HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. Methods Inf Med 2018 Jul;57(S 01):e66-e81.
20. Winter A, Stäubert S, Ammon D, Aiche S, Beyan O, Bischoff V, et al. Smart Medical Information Technology for Healthcare (SMITH). Methods Inf Med 2018 Jul;57(S 01):e92-e105.
21. Prasser F, Kohlbacher O, Mansmann U, Bauer B, Kuhn KA. Data Integration for Future Medicine (DIFUTURE). Methods Inf Med 2018 Jul;57(S 01):e57-e65.
22. AI for humanity. [Online]. Available: https://www.aiforhumanity.fr. [Accessed: 05-May-2019].
23. Discours du Président de la République sur l'intelligence artificielle [Speech by the President of the French Republic on artificial intelligence]. elysee.fr. [Online]. Available: https://www.elysee.fr/emmanuel-macron/2018/03/29/discours-du-president-de-la-republique-sur-lintelligence-artificielle.

[Accessed: 05-May-2019].
24. Tuppin P, Rudant J, Constantinou P, Gastal-di-Ménager C, Rachas A, de Roquefeuil L, et al. Value of a national administrative database to guide public decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. Rev Epidemiol Sante Publique 2017 Oct;65 Suppl 4:S149-S167.
25. Bezin J, Duong M, Lassalle R, Droz C, Pariente A, Blin P, et al. The national healthcare system claims databases in France, SNIIRAM and EGB: Powerful tools for pharmacoepidemiology. Pharmacoepidemiol Drug Saf 2017 Aug;26(8):954-62.
26. LOINC — The freely available standard for identifying health measurements, observations, and documents. [Online]. Available: https://loinc.org/. [Accessed: 05-May-2019].
27. Kalra D, Beale T, Heard S. The openEHR Foundation. Stud Health Technol Inform 2005;115:153–73.
28. Index - FHIR v4.0.0. [Online]. Available: http://www.hl7.org/FHIR/. [Accessed: 05-May-2019].
29. Common Terminology Services 2TM (CTS2TM) | Object Management Group. [Online]. Available: https://www.omg.org/cts2/index.htm. [Accessed: 05-May-2019].
30. MIRACOLIX Tools – MIRACUM. [Online]. Available: https://www.miracum.org/miracolix-tools/. [Accessed: 05-May-2019].
31. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc 2010 Mar-Apr;17(2):124-30.
32. Athey BD, Braxenthaler M, Haas M, Guo Y. tranSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research. AMIA Jt Summits Transl Sci Proc 2013 Mar 18;2013:6-8.
33. da Silva Santos LB, Wilkinson MD, Kuzniar A, Kaliyaperumal R, Thompson M, Dumontier M et al. FAIR Data Points Supporting Big Data Interoperability. Enterprise Interoperability in the Digitized and Networked Factory of the Future. London: ISTE; 2016. p. 270–9.
34. Dixon BE, McGowan JJ, Grannis SJ. Electronic laboratory data quality and the value of a health information exchange to support public health reporting processes. AMIA Annu Symp Proc 2011;2011:322-30.
35. Duarte J, Castro S, Santos M, Abelha A, Machado J. Improving Quality of Electronic Health Records with SNOMED. Procedia Technol 2014;16:1342–50.
36. Daniel C, Serre P, Orlova N, Bréant S, Paris N, Griffon N. Initializing a hospital-wide data quality program. The AP-HP experience. Comput Methods Programs Biomed 2018 Nov 9.
37. Woinarowicz M, Howell M. The impact of electronic health record (EHR) interoperability on immunization information system (IIS) data quality. Online J Public Health Inform 2016;8(2):e184.
38. Dixon BE, Vreeman DJ, Grannis SJ. The long road to semantic interoperability in support of public health: Experiences from two states. J Biomed Inform 2014 Jun;49:3-8.
39. Systèmes d'information de santé : le diagnostic est posé, le traitement s'impose [Health information systems: the diagnosis is made, treatment is needed]. [Online]. Available: https://www.senat.fr/rap/r07-035/r07-03511.html. [Accessed: 05-May-2019].
40. Cadre d'Interopérabilité des Systèmes d'Information de Santé [Framework for Interoperability of Health Information Systems]. [Online]. Available: https://esante.gouv.fr/interoperabilite/ci-sis. [Accessed: 05-May-2019].
41. Lee D, de Keizer N, Lau F, Cornet R. Literature review of SNOMED CT use. J Am Med Inform Assoc 2014;21(e1):e11–e19.
42. Séroussi B, Bouaud J. Adoption of a Nationwide Shared Medical Record in France: Lessons Learnt after 5 Years of Deployment AMIA Annu Symp Proc 2017;2016:1100–9.
43. Ministère des Solidarités et de la Santé. Groupements hospitaliers de territoire. Ministère des Solidarités et de la Santé [Ministry of Solidarities and Health. Territory Hospital networks. Ministry of Solidarity and Health], 05-May-2019. [Online]. Available: https://solidarites-sante.gouv.fr/professionnels/gerer-un-etablissement-de-sante-medico-social/groupements-hospitaliers-de-territoire/. [Accessed: 05-May-2019].
44. Mise en place d'un entrepôt de données de santé et de son cadre d'utilisation [Deployement of a health data warehouse and its framework of uses]. [Online]. Available: /contenu/mise-en-place-dun-entrepot-de-donnees-de-sante-et-de-son-cadre-dutilisation. [Accessed: 05-May-2019].
45. Daniel C, Serre P, Orlova N, Bréant S, Paris N, Griffon N. Initializing a hospital-wide data quality program. The AP-HP experience. Comput Methods Programs Biomed 2018 Nov 9.
46. Heudel P, Livartowski A, Arveux P, Willm E, Jamain C. [The ConSoRe project supports the implementation of big data in oncology]. Bull Cancer 2016 Nov;103(11):949-50.
47. Madec J, Bouzille G, Riou C, Van Hille P, Merrour C, Delamarre D, Raimbert V, et al. eHOP Clinical Data Warehouse : from a prototype to the creation of an inter-regional clinical data centers network. Medinfo 2019 - Accepted.
48. Bialke M, Bahls T, Havemann C, Piegsa J, Weitmann K, Wegner T, et al. MOSAIC--A Modular Approach to Data Management in Epidemiological Studies. Methods Inf Med 2015;54(4):364-71.
49. CNIL. [Online]. Available: https://www.cnil.fr/. [Accessed: 05-May-2019].
50. Arrêté du 5 mai 2017 portant nomination au Comité d'expertise pour les recherches, les études et les évaluations dans le domaine de la santé [Order of May 5, 2017 appointing the Expert Committee for Research, Studies and Evaluations in the field of health]. Legifrance [WWW Document], n.d. URL https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000034680198&categorieLien=id. [Accessed 17-June 2019].
51. subsidiarity | Definition of subsidiarity in English by Oxford Dictionaries. Oxford Dictionaries | English. [Online]. Available: https://en.oxforddictionaries.com/definition/subsidiarity. [Accessed: 05-May-2019].
52. Swiss Personalized Health Network (SPHN) named a 2019 GA4GH Driver Project. [Online]. Available: https://dcc.sib.swiss/news/ga4gh-driver/. [Accessed: 05-May-2019].
53. The future of healthcare: our vision for digital, data and technology in health and care. GOV.UK. [Online]. Available: https://www.gov.uk/government/publications/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care. [Accessed: 05-May-2019].
54. Code of conduct for data-driven health and care technology. GOV.UK. [Online]. Available: https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology. [Accessed: 05-May-2019].
55. Le Réseau santé wallon devient l'interface d'un portail de traitement des données médicales [The Walloon Health Network becomes the interface of a medical data processing portal]. 19-Jul-2018. [Online]. Available: https://www.lejournaldumedecin.com/actualite/le-reseau-sante-wallon-devient-l-interface-d-un-portail-de-traitement-des-donnees-medicales/article-normal-35161.html. [Accessed: 05-May-2019].
56. Kalra D, Stroetmann V, Sundgren M, Dupont D, Schlünder I, Thienpont G, et al. The European Institute for Innovation through Health Data. Learn Health Syst 2017;1(1):e10008.
57. Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. J Biomed Inform 2017 Sep;73:14-29.
58. PARS3 : Plateforme apprenante pour la recherche en santé et services sociaux au Québec [PARS3 : Learning platform for research in health and social services in Quebec]. GRIIS. [Online]. Available: http://griis.ca/pars3/. [Accessed: 05-May-2019].
59 INITIATIVE CLARET. GRIIS. [Online]. Available: https://griis.ca/initiative-claret/. [Accessed: 05-May-2019].
60. International comparison of digital strategies. [Online]. Available: https://www.bertelsmann-stiftung.de/en/publications/publication/did/smarthealthsystems-1/. [Accessed: 26-May-2019].

Correspondence to:
Marc Cuggia
Equipe DMS/Inserm LTSI, Faculté de Médecine
Rue du Pr Léon Bernard
35043 Rennes
France
E-mail: marc.cuggia@univ-rennes1.fr