NATIONAL CONSULTATIVE ETHICS COMMITTEE
FOR HEALTH AND LIFE SCIENCES

NATIONAL PILOT COMMITTEE
FOR DIGITAL ETHICS

under the *aegis of*
NATIONAL CONSULTATIVE ETHICS COMMITTEE
FOR HEALTH AND LIFE SCIENCES

JOINT OPINION
CCNE OPINION 143/CNPEN OPINION 5

HEALTH DATA PLATFORMS: ETHICAL ISSUES

Opinion adopted unanimously by the members present at the CCNE plenary session on 16 February 2023

Opinion adopted unanimously by the members present at the CNPEN plenary session on 28 February 2023

Citation of this opinion:
Health data platforms: ethical issues. Joint opinion of the CCNE and the CNPEN, Opinion 143 of the CCNE, Opinion 5 of the CNPEN. February 2023.

CONTENTS

SUMMARY

The proliferation of public and private health data collection operations and the complications associated with accessing them have highlighted the importance of this data, but also the tensions and fears that its use raises. Increasingly, this health data is being gathered together in digital infrastructures, known as health data platforms (HDPs), which also offer access and processing tools.

The vast landscape of these private and public platforms, and their growing development in a context that is currently largely unregulated, means that a global analysis is needed to consider the consequences of decisions relating to the collection, processing and use of this sensitive information. In addition, the hardware and software architecture as well as the organisation and human resources devoted to such platforms need to be examined as a whole.

In order to inform decisions and public policies relating to the design and implementation of HDPs, the National Consultative Ethics Committee (CCNE) and the National Pilot Committee for Digital Ethics (CNPEN) have taken it upon themselves to conduct a joint study that takes account of the issues involved in both health ethics and digital ethics. Members of the regional ethics committees (ERER) were also involved in the discussions.

The CCNE and the CNPEN have developed their reflections by firstly endeavouring to provide as exhaustive a definition as possible of what health data is, and to develop, through concrete examples, what its usefulness and possible uses are. The committees emphasise that health data is not a commodity, but a personal attribute, and therefore cannot be traded unless anonymised, bearing in mind that no anonymisation process is currently certified. A typology of infrastructures is then proposed in order to clarify the current HDP landscape by identifying the operational scope of these infrastructures and the ethical issues underlying technical choices and innovations. The opinion then looks at the issues surrounding sovereignty, paying particular attention to the multiple meanings of the term, which brings into conflict several perspectives: liberal and entrepreneurial, regulatory and protective, and finally an alternative approach known as strategic autonomy. These clarifications of the concept of sovereignty highlight the ethical tensions involved, based on the principles of beneficence, justice, equity in healthcare systems, and explicability and transparency. The discussion then turns to the valuation of health data, identifying two different economic models that raise distinct ethical issues.

Finally, the last part of the opinion is devoted to the different types of consent to the use of health data, in particular the default strategy and altruism with regard to health data, and to citizen participation in the governance of HDPs. It appears that new forms of dynamic consent are needed insofar as the data stored in the platforms is likely to be used for purposes other than that for which the individual initially gave consent. The CCNE and CNPEN are particularly attentive to issues relating to citizen participation in the construction of health data infrastructures and their governance. Numerous surveys on this subject show that the public is not particularly attentive to these issues if they are not relayed by patient associations, which play a very important role in this area.

In the course of this opinion, the CCNE and the CNPEN put forward 21 recommendations (3 of which relate more specifically to research and innovation), which are grouped together at the end according to the themes they cover: the quality and sharing of health data (2 recommendations), the environmental impact of HDPs (1), their architecture (4), the anonymisation of data (1), sovereignty (4), the valuation of data (3), and the conditions for a collaborative ecosystem for HDPs (6).

INTRODUCTION

1. Background to the self-referral

In May 2019, the National Consultative Ethics Committee for Health and Life Sciences (CCNE) highlighted that "the mass accumulation of data derived from individuals, and the increased capacity in the processing of this data to produce value, require debate and ethical reflection"[1]. The crisis caused by the Covid-19 pandemic highlighted the importance of collecting and accessing health data, but also the tensions, misgivings and fears that its use arouses. Increasingly, this health data is being gathered together in digital infrastructures, known as health data platforms (HDPs), which also offer access and processing tools.

The growing development of private and public platforms collecting health data - gathered by laboratories, hospitals, clinics, general practitioners and other players outside the healthcare system - calls for a global analysis of the long-term consequences of decisions relating to the collection, sharing, preservation, processing and use of this sensitive information. The creation of the *Groupement d'Intérêt Public - Plateforme des données de santé* (GIP-PDS), commonly known as the *Health Data Hub*[2], did not bring these reflections to a close. Furthermore, the hardware and software architecture as well as the organisation and human resources devoted to such platforms need to be examined as a whole.

These collections of personal health data raise technical issues (storage, security, anonymisation/pseudonymisation, standardisation, sharing, etc.), legal issues (status of the data, ownership regime to be adopted, consent, etc.), but the questions and debates they raise, while crucial, must not overshadow the underlying ethical issues.

In order to inform decisions and public policies relating to the design and implementation of HDPs, the National Consultative Ethics Committee (CCNE) and the National Pilot Committee for Digital Ethics (CNPEN) decided to carry out this joint study to take account of issues relating to both health ethics and digital ethics. Members of the Espaces de Réflexion Éthique Régionaux (ERER) (regional ethics committees) were also involved in the discussions.

Since health data is a vital intangible asset, its availability and digital use must not run counter to the fundamental rights of individuals. A balance must be struck between the requirements of the public interest and those guaranteeing respect for privacy.

The construction of health databases, cohorts, warehouses and HDPs represents a major investment, often financed out of the public purse. It is vital that the values upon which we want to continue to build our healthcare system be clearly defined. This means defining the legal status of this data and the means of compensating the various players who have contributed to the construction, maintenance and use of these databases, cohorts, warehouses and HDPs. The articulation and harmonisation of rules for securing and accessing information must also be anticipated.

The creation of a one-stop shop centralising a vast amount of information, on the one hand, and the networking of multiple independent and specialised platforms, on the other, are solutions that each have their advantages and disadvantages in terms of governance,

---

1 CCNE, Opinion 130, 29 May 2019, *Données massives et santé : une nouvelle approche des enjeux éthiques*, 94 p.

2 Paris Administrative Court ruling of 22 October 2022 concerning the name of the GIP-PDS.

security, financial cost, potential value creation and the computing capacity required to process the information.

Finally, over and above individual consent to making personal health data available to individuals in general, and patients in particular, there is a particular health democracy issue at stake, which raises questions about citizen participation in the governance of HDPs and the role of patients' or carers' associations in the development of research projects.

We must consider how these strategic choices fit into the broader picture of our healthcare system and its values.

It is important for France to have a clear policy, so that it can both claim real sovereignty over its health data and contribute to European and international efforts in the field of public health. It must therefore equip itself with sufficient technical resources to be able to process and analyse this information at national and European level.

In this way, it will be able to consider and shape the way in which these new resources fit into its healthcare system and organise the sharing of their potential benefits, while respecting the values of solidarity, human dignity, justice and autonomy that it embodies. An international comparison in this area may be useful in assessing the strengths, identifying the difficulties to be overcome and the partnerships to be created.

Box 1 - Health Data, Health Databases, Health Data Platforms

Various concepts are used in this opinion, and we feel it is important to define them from the outset.

Health data: personal data relating to the health (physical or mental) of patients, collected by laboratories, hospitals, clinics, general practitioners or other parties involved in the healthcare process.

Health database: a structured and organised set of data enabling large quantities of information relating to a specific area of health to be stored and used.

Health data platform: private or public digital infrastructures providing access to and processing of health data.

2. Bioethics and digital ethics issues

The joint reflections of the CCNE and the CNPEN have been informed by the shared and specific values of bioethics and digital ethics. The reflections of the various contributors to the work *Pour une éthique du numérique*[3] shed light on this subject.

Box 2: CNPEN manifesto "For digital ethics".

The manifesto of the National Pilot Committee for Digital Ethics (CNPEN), drafted at its annual seminar (held jointly with the CCNE) on 15 and 16 September 2020 and published in April 2021, identifies the foundations for reflection on digital ethics.

In particular, it takes into account the systematic quantification and evaluation of human activities, which raises questions about our relationship to knowledge and memory. Digital

[3] National Pilot Committee for Digital Ethics - *Pour une éthique du numérique*. É Germain, Cl. Kirchner, C. Tessier, PUF 2022, ISBN 978-2-13-083348-2.

ethics is thus challenged to take into account the way in which we consider human autonomy. The manifesto also points out that these technologies and the economic models that support them are overturning the various areas of sovereignty. Ethical reflection led by a committee is therefore essential for the people and institutions that develop, market, regulate and use digital technologies.

There is considerable overlap between the work of the CCNE and the CNPEN. Bioethics and digital ethics (or cyberethics) share a number of principles: respect for human dignity and autonomy, nonmaleficence, equity and justice. These two ethical reflection practices overlap in many respects: they are based on collective, multi-disciplinary reflection, focusing on the evaluation of scientific innovations in their context of use; they identify the tensions that these innovations raise between these principles and seek to evaluate the consequences that can reasonably be attributed to them.

However, each of these two fields of ethical reflection has its own distinctive aspects. Most authors agree on this point, even if they do not all point to the same differences, which relate first and foremost to the principles governing each activity. For example, while bioethics requires medical practice to aim for the good of the patient, no one is asking the developers of digital tools to adhere to such a requirement[4]. These tools are designed to meet the needs of users, which are developed to varying degrees, as part of an economic, social or public management activity. This leads us to see digital ethics as a first step towards the development of regulations, which are still too fragmented, aimed at protecting the uses of digital technology, whereas bioethics already has well-established legal support.

From another angle[5], however, we might consider that the difference between bioethics and cyberethics stems from the fact that, while healthcare practices are primarily about human beings, digital tools primarily manipulate *information*[6] in large quantities and of a heterogeneous nature, which raises unique problems. In particular, digital technology raises issues of information overload or 'infobesity', which can lead to misinformation; it also makes it possible for the public sphere to intrude into the private or family sphere (as in the case of harassment). This makes it necessary to assess the quality and robustness of tools (which is also an issue in medicine, but where it is much better regulated); and lastly, digital technology raises questions about the transparency of systems, formulated in cyberethics in terms of explicability (i.e. the operation of a tool must be presented in a way that can be understood by a reasonably literate person). This issue also arises in bioethics, but is a more pressing challenge in digital ethics with the advent of machine learning algorithms whose operation is often difficult to understand and which are sometimes described as opaque.

Faced with these specific aspects of digital ethics, this opinion proposes to take a maximalist approach: it will take into account not only the principles common to both types of ethical reflection, but also those that are specific to one or the other.

---

4 R. Chatila - Chapter "*Bioéthique et éthique du numérique : une hybridation paradoxale*" in *Pour une éthique du numérique*, *op. cit.*

5 C. Froidevaux, G. Adda, Chapter "*Regards croisés sur la cyberéthique et la bioéthique*", in *Pour une éthique du numérique*, *op. cit.*

6 Healthcare professionals also handle information, but it is collected with a defined objective (to improve the patient's health), whereas digital professionals handle very large quantities of highly heterogeneous data collected for various purposes and reused for other purposes, some of which are health-related.

Principles of biomedical ethics	Principles of digital ethics
• **Principle of autonomy**: obligation to respect the decision-making capacity and consent of autonomous individuals;	• **Principle of autonomy**: preserve the human capacity to act on tools and data;
• **Principle of beneficence**: the obligation to provide benefits and to weigh up the benefits against the risks;	• *While the patient is at the heart of biomedical ethics, not all digital systems are designed for the good of their users.*
• **Principle of nonmaleficence:** obligation to avoid harm;	• **Principle of nonmaleficence**: do no harm nor exacerbate harm (safety, security, technical robustness);
• **Principle of justice**: obligation of equity, non-discrimination, fair distribution of benefits and risks.	• **Principle of justice**: equity, reduction of bias, non-discrimination, proportionality;
• *The principle of explicability is present in medical practice in connection with informed consent.*	• **Principle of explicability**: transparency, interpretability, traceability, auditability. A fundamental principle with the advent of deep learning.
According to the book *Pour une éthique du numérique*[7] which is based on the one hand on Beauchamp's principles of bioethics[8],on the other hand on the report of the *High-level expert group on artificial intelligence* set up by the European Commission in June 2018[9].	

The French Council of State's recent report to the Prime Minister on Artificial Intelligence (AI) and public action[10] proposed the adoption of seven principles that correspond to those in the table above, but with a more detailed definition:

- Human primacy in several aspects: benefits, supervision, non-dependence and assistance, social acceptability of "machine error", indirect form of human error;
- Performance: indicators (accuracy, response time, satisfaction rate), acceptable levels and determining factors of this performance;
- Equity and non-discrimination: type of equity, risk of algorithmic bias, accessibility and universality;

[7] R. Chatila - Chapter "*Bioéthique et éthique du numérique : une hybridation paradoxale*" in *Pour une éthique du numérique*, *op. cit.*, p.34

[8] Beauchamp T.L., Childress J., (1979, 1st edition), *Principles of biomedical ethics*, New York: Oxford University 314 p. ; Beauchamp T.L., (2003), *Methods and principles in biomedical ethics*, J Med Ethics, Oct;29(5):269-74. doi: 10.1136/jme.29.5.269. PMID: 14519835; PMCID: PMC1733784.

[9] High-level expert group on artificial intelligence,European Commission, *Ethical guidelines for trustworthy AI*, April 2019 - See: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[10] Council of State, *Intelligence artificielle et action publique : construire la confiance, servir la performance*, Study commissioned by the Prime Minister, 31/03/2022, https://www.conseil-etat.fr/publications-colloques/etudes/intelligence-artificielle-et-action-publique-construire-la-confiance-servir-la-performance.

- Transparency: right of access to system documentation, requirement for fairness, explicability, transparent design and auditability;
- Security (cybersecurity);
- Environmental sustainability;
- Strategic autonomy.

Box 3: CCNE Opinion 130 "Données massives et santé : une nouvelle approche des enjeux éthiques" [Mass data and health: a new approach to the ethical issues]"

In response to a referral from the Minister for Social Affairs and Health concerning the ethical issues associated with the collection and processing of mass data in the field of health, the CCNE published its opinion 130 entitled "Mass data and health: a new approach to the ethical issues" in May 2019. In this opinion, which introduces the ethical issues raised by the complexity of the digital revolution, the CCNE takes stock of mass data in the field of health and proposes an analysis of the ethical issues that have emerged with the development of the collection, processing and digital exploitation of health data. Readers may wish to refer to this opinion for a more in-depth analysis of the diversity of stakeholders, data and their objectives, and the implications of these far-reaching changes for the protection of personal data.
Three of the recommendations in CCNE opinion 130 are of particular relevance to this opinion:
Recommendation 10: the CCNE recommends the development of national mutual and interconnected platforms;
Recommendation 11: the CCNE considers that the ethical imperative in research must be adapted to each specific situation, so as to justify a relationship of trust between the holders of the data and those who have access to it and process it;
Recommendation 12: the CCNE considers that it is necessary to facilitate the sharing of health data for research purposes.

I. FROM HEALTH DATA TO HEALTH DATA PLATFORMS

1. Personal health data is not a commodity

1.1 Principle of non-transferability of personal health data

The General Data Protection Regulation (GDPR) of 27 April 2016, applied since May 2018, gives a somewhat broad definition of health data (Article 4. 15 and recital 35). We repeat here the presentation of the French Data Protection Authority (CNIL) in its note "*Qu'est-ce qu'une donnée de santé ?*" [What is health data?] [11]:

> "Personal data concerning health is data relating to the past, present or future physical or mental health of a natural person (including the provision of healthcare services) which reveals information about that person's state of health. This therefore includes information relating to a natural person collected at the time of his or her registration with a view to receiving healthcare services or during the provision of such services [...]); information obtained during the testing or examination of a body part or bodily substance, including from genetic data and biological samples; information relating to a disease, disability, risk of disease, medical history, clinical treatment or the physiological or biomedical state of the person concerned (...). This definition makes it possible to include certain measurement data from which it is possible to deduce information about the individual's state of health".

Health data is thus defined by its purpose.

In French law, as in European law, **personal data is not linked to property rights but to personality rights**. The GDPR has recently further enshrined this concept inspired by personalism[12]. Databases, on the other hand, are protected by copyright or intellectual property law.

Personal data is an item of information covered by freedom of expression, which means that it cannot be appropriated or transferred. **Classified as a personal attribute**, and therefore covered by personality rights, it is very closely linked to private life[13]. In the case of personal health data in particular, it concerns the most intimate aspect of the functioning of the human body[14]; as such, it is surrounded by more guarantees than other data, particularly with regard to its processing. Individuals enjoy very strong protection, which in some ways runs counter to their freedom, since they are prohibited from selling their body, or an organ of their body, or information relating to that body. This is why Article 1111-8 of the French Public Health Code prohibits, on pain of criminal sanction, "any act of transfer for valuable consideration of identifying health data, directly or indirectly, including with the consent of the person concerned".

However, the protection of personal data, which has recently been strengthened in French law and then in European Union law, is not only defensive and protective of the individual against itself and the institutions, but also gives the individual an active role. For example,

---

[11] CNIL, "Qu'est-ce qu'une donnée de santé ?" [https://www.cnil.fr/fr/quest-ce-ce-quune-donnee-de-sante].

[12] Inspired by Kantian (Charles Renouvier) and spiritualist (Nicolas Berdaiev, Emmanuel Mounier) ideas, personalism makes the human person - a rational, voluntary and relational being, as opposed to an egocentric individual - the centre and foundation of all possible knowledge, creativity and involvement in the world. See https://fr.wikipedia.org/wiki/Personnalisme

[13] Article 9 of the French Civil Code: "Everyone has the right to respect for his or her private life".

[14] Article 16-1 of the French Civil Code:: "Everyone has the right to respect for his or her body; the human body is inviolable. The human body, its components and products may not be the subject of property rights". Article 16-5: "agreements that have the effect of conferring a pecuniary value on the human body, its parts or its products are null and void".

the law of 7 October 2016 for a Digital Republic introduced a second paragraph into Article 1 of the law of 6 January 1978 recognising every person's "**right to decide on and control** the uses made of personal data concerning them". This development can be seen in particular in the creation of new rights, such as the possibility of giving instructions on the fate of the person's data after death, the right to be forgotten and the right "**to data portability**".

1.2 Different uses and origins of health data

Health data is the responsibility of an individual, but also of the health professional who examined that individual, and of the person who interprets the data. This means that health data is enriched by parties other than the person to whom it refers.

In its note on health data[15], the CNIL distinguishes three categories of data: (i) health data by its very nature, such as that relating to diseases, (ii) data that becomes health data after cross-referencing with other data, insofar as it enables a conclusion to be drawn about the state of health or health risk of an individual, and (iii) data whose medical use makes it health data. The first category is also referred to as health data for a primary use and the second and third categories as health data for secondary uses.

Health data is very specific and sensitive personal data. Recital 51 of the GDPR states that: "Personal data which are, by their nature, particularly sensitive in relation to fundamental rights and freedoms merit specific protection as the context of their processing could create significant risks to the fundamental rights and freedoms".

Under Article 9 of the GDPR, the processing of health data is in principle prohibited unless the data subject has given his or her consent. It is also permitted even if consent has not been given, in a certain number of cases, in particular the preservation of the vital interests of the data subject, if he or she is unable to give consent, and the needs of the management of health or social protection systems and services, occupational medicine, preventive medicine, diagnosis, care and treatment[16].

It should be noted that in France, the use of health data for research involving human subjects is governed by a relatively complex validation cycle, which is presented in the appendix (Appendix 5).

In addition to this health data obtained in a medical context, other data can be obtained by sensors in connected objects such as connected watches worn by people, whether sick or healthy, which are outside the scope of medical diagnosis in the true sense of the term, but which can provide a continuous record of heart rate, electrocardiogram, perspiration, stress levels, sleep, etc., in a wide variety of living conditions: rest, work, sport. These are known as 'wellness applications', but the data is often relevant and complementary to health data in the strict sense of the term.

The possibility of including them in HDPs, and in particular in the personalised digital medical file, as in *Mon espace santé*, is therefore tempting, but poses specific problems of access rights for the individuals concerned and for healthcare professionals, and of making them available to third parties by the digital platforms that store them and possibly process them using algorithms to derive personalised diagnoses or public health statistics. In this regard, *Mon espace santé* has a procedure for listing the applications available to users,

---

15 CNIL, "*Qu'est-ce qu'une donnée de santé ?*" [https://www.cnil.fr/fr/quest-ce-ce-quune-donnee-de-sante]

16 CNIL, "*Recherche médicale : quel est le cadre légal ?*" , [https://www.cnil.fr/fr/quelles-formalites-pour-les-traitements-de-donnees-de-sante-caractere-personnel]

but does not choose between competing applications. This choice is therefore left to the user, who may feel helpless when faced with an overabundance of offers.

The recent CCNE and CNPEN opinion on medical diagnosis and AI[17] states: "More subjective data can also be collected using questionnaires incorporated into smartphone applications or by recording behavioural images [...] While such data can help to identify people requiring care more easily and quickly, the processing or dissemination of this information can also expose them to significant risks and raise important ethical questions".

1.3 Data relating to specific ethnic characteristics

Among health data, it is possible to single out that which relates to the individual's real or supposed ethnicity. Some diseases particularly affect certain populations, such as sickle cell anaemia, which mainly affects people from sub-Saharan Africa and the Caribbean. It is therefore essential to have information on ethnicity in order to advance research, otherwise certain diseases may be overlooked. However, France has a very complex relationship with ethnic data, which is specific to it in contrast to other European countries[18].

Today, the French regulatory framework, based on Article 1 of the Constitution and Article 6.1 of the Data Protection Act[19], prohibits the collection of ethnic data, except for a specific purpose with special authorisation from the CNIL. One example is the TeO (Trajectoires et Origines) survey conducted by the French National Institute for Demographic Studies (INED) on the issue of population diversity in France[20]. In the field of medical research, the CNIL authorises "research requiring an examination of genetic characteristics" as part of its 2018 reference methodologies, subject to certain conditions[21]. However, this focus on exceptions based on their purpose does not resolve the problem posed by the possible availability of this data in HDPs, as the uses to which it may subsequently be put are not known in advance.

In this regard, the minutes of the "Information mission on the emergence and development of the various forms of racism and the responses to be made to them"[22] stress that:

> "The French Data Protection Act and the GDPR provide guarantees with regard to so-called "sensitive" data, including data revealing racial or ethnic origin. The recitals of the GDPR state that "the use of the term 'racial origin' in this Regulation does not imply an acceptance by the [European] Union of theories which attempt to determine the existence of separate human races". French law refers to the "supposed racial origin of individuals".

This issue is so complex, even if we restrict it to health data alone - so much so that it is difficult to choose the vocabulary to talk about it, as the title of this sub-section shows - that it cannot be dealt with in the present framework of our reflection on HDPs. We believe

---

[17] Joint opinion no. 141 of the CCNE and no. 4 of the CNPEN, (2023), *Diagnostic Médical et Intelligence Artificielle : Enjeux Ethiques*, 58 p.

[18] Le Monde, "*Statistiques ethniques: une situation contrastée en Europe*", 05/02/2010.

[19] See: https://www.cnil.fr/fr/la-loi-informatique-et-libertes#article6: "It is prohibited to process personal data revealing the supposed racial or ethnic origin [...] of a natural person or to process genetic data, biometric data for the purpose of identifying a natural person in a unique way [...]"

[20] See the survey website: https://teo1.site.ined.fr/.

[21] CNIL, "*Méthodologie de référence MR-001. Recherches dans le domaine de la santé avec recueil du consentement*" [https://www.cnil.fr/fr/declaration/mr-001-recherches-dans-le-domaine-de-la-sante-avec-recueil-du-consentement].

[22] *Mission d'information sur l'émergence et l'évolution des différentes formes de racisme et les réponses à y apporter, Compte rendu n° 47*, 17 November 2020. [https://www.assemblee-nationale.fr/dyn/15/comptes-rendus/racisme/l15racisme2021047_compte-rendu#].

that this is a profound and serious issue. It has been included in the work programme of the CCNE and the CNPEN, which will produce an opinion on the matter within a reasonable timeframe.

2. Classification of health data infrastructures

In this opinion, we are interested in the structures that collect data or make it possible to exchange this data, and offer the means to process this information. Organisations that collect health data may be private or public: they may be laboratories, hospitals, clinics, or sometimes players on the fringes of the care pathway. Health data structures are called health information systems or health databases (HDB), health data warehouses (HDW) or health data platforms (HDP). They may be local, with a single physical storage location (hub), organised as a network or interrogated via a mediating platform. The most sophisticated ones offer processing resources (computing space, numerical processing software, AI algorithms, etc.). We begin by defining these different types of health data structures, before focusing on health data platforms.

2.1 Health databases

With regard to **health databases (HDB)**, we shall take as our starting point the definition provided in CCNE opinion 130[23], which we shall explain in greater detail: a database is a structured and organised collection enabling large quantities of information relating to a specialised field to be stored for use. This definition is in line with Article 1 of Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, which are seen as a "collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means". Our definition emphasises the large-scale nature of databases. Data can be integrated into a database as it is (possibly supplied by various external sources) or annotated (processed by the managers of these databases). The main uses of a database are access to the medium for retrieving, exploiting and disseminating data, and querying for statistical purposes, for example.
Some health databases are derived from the PMSI[24] (Programme de Médicalisation des Systèmes d'Information), which is used to "describe the medical activity of health establishments in a synthetic and standardised way. It is based on the recording of standardised medico-administrative data in a standard data collection system". This is the case, for example, with the hospital databases that collect and structure data on follow-up and rehabilitation care (PMSI-SSR[25]).

2.2 Health data warehouses

In the field of information technology, a data warehouse is an IT infrastructure that brings together data in a single physical location. This data can be expressed in a variety of formats and comes from a number of sources that are often heterogeneous and sometimes of very different natures. The physical nature of the warehouse means that it is a physical location comprising hardware and human resources. Generally speaking, the IT

[23] CCNE, Opinion 130, 29 May 2019, *Données massives et santé : une nouvelle approche des enjeux éthiques*, p. 64.

[24] See: https://solidarites-sante.gouv.fr/professionnels/gerer-un-etablissement-de-sante-medico-social/financement/financement-des-etablissements-de-sante-10795/financement-des-etablissements-de-sante-glossaire/article/programme-de-medicalisation-des-systemes-d-information-pmsi

[25] See: https://www.atih.sante.fr/presentation-pmsi-ssr

data warehouse enables this heterogeneous data to be integrated into a unified model, making it easier for the user to exploit. It should be noted that the term **health data warehouse (HDW)** is used even in the absence of a unified model enabling effective IT integration, as long as the link between the different data in the databases brought together in the warehouse can be established. In addition, HDWs are increasingly coupled with data platforms (defined below) to enable *in situ* processing of their data.

In 2019, the CNIL made a formal distinction[26] between an HDB, a simple IT structure that collects health data for the purpose of a one-off research, study or evaluation, and an HDW, an IT structure that enables several processing operations to be carried out at a later date. On 17 November 2021, the CNIL adopted a data warehouse standard[27], designed to simplify procedures. This standard means that organisations wishing to implement an HDW that complies with the standard do not need to seek prior authorisation from the CNIL. It applies only to HDWs based on the performance of a public-interest mission, as defined in Article 6.1.e of the GDPR. Under this standard, it considers that "HDWs are databases intended to be used in particular for the purposes of research, studies or evaluations in the field of healthcare". In this regard, it should be noted that the re-use of data from an HDW for a specific project constitutes "data processing" in its own right within the meaning of the GDPR; each project must therefore have its own data processing area, separate from other projects on the HDW.

2.3 Health data platforms

We propose defining a **health data platform** (HDP) as an HDW that also offers services for sharing, processing and analysing data, such as software and computing capacity on high-capacity servers. Algorithmic processing may be based on machine learning, and very often requires large amounts of data. Without claiming to be exhaustive, various types of French or foreign HDPs are presented in the appendices (appendices 4.1 to 4.8), ranging from long-established HDBs and HDWs that have equipped themselves with services, to structures created recently to pool and increase the range of services on offer, such as the GIP-PDS.

2.4 Cohorts

A cohort is a particular type of HDB, and is one of the principal tools used in epidemiology to study the distribution of disease and disability in human populations, and the influences that determine this spread. It involves selecting a group of volunteers who may share a certain number of common characteristics, and tracking them over time at the individual level in order to identify the occurrence of health events of interest. In France, there were more than 250 cohort studies in 2019. However, some cohorts involving tens or hundreds of thousands of people can be made available to several research projects and thus become veritable health data platforms. Two examples of this are provided in the appendices (Appendix 4.9 and 4.10): Constances, based on a French cohort, and UK Biobank, based on a British cohort.

---

[26] CNIL, "*Traitements de données de santé : comment faire la distinction entre un entrepôt et une recherche et quelles conséquences ?*" [https://www.cnil.fr/fr/traitements-de-donnees-de-sante-comment-faire-la-distinction-entre-un-entrepot-et-une-recherche-et]

[27] CNIL, "*La CNIL adopte un référentiel sur les entrepôts de données de santé*" [https://www.cnil.fr/fr/la-cnil-adopte-un-referentiel-sur-les-entrepots-de-donnees-de-sante]

2.5 Data brokers

Alongside health data platforms, companies have emerged that collect personal information, typically through online activities, in order to organise their market: these are known as **data brokers**. This commercialised data may be either primary-use health data, but in principle anonymised (see section I.3.2.2), or personal information derived from traces left on the internet, such as messages on social networks or forums, but also through health applications (IoT[28], trackers, sensors) or well-being applications (fitness, sport). This can also include order receipts from online pharmacies, online consultation histories and other sources of public or non-public medical information. In addition, data on the location of and visits to medical facilities (clinics, hospitals) or gyms can be included. It should be noted that it can be difficult to trace the origin of this data. One example of a health data broker is IQVIA-France[29], which specialises in the drugs market and has been accumulating data relating to antigen tests and anti-Covid vaccines since the start of the pandemic. Another example is Cegedim[30], which develops and markets health databases and software.

2.6 Data exchange platforms

In addition to data brokers, whose main purpose is to market access to collected data, other commercial operators worth mentioning include data exchange platforms, which bring data suppliers and purchasers into contact by enabling them to exchange data in a secure environment, without ever storing the data. These exchange platforms may be administered by companies, such as Dawex, or cooperatives, such as Salus Co-op in Spain, an open-access data exchange platform, Healthbank, a pay-as-you-go initiative in Switzerland, or Doctolib in France. These examples of health data exchange platforms are presented in the appendices (Appendices 4.11 to 4.14). Several of them would be eligible for the status of "data intermediation services provider" under the European Data Governance Act of 23 June 2022[31] (see section III.4.2).

3. Why exchange, collect and process health data on a mass scale?

The reasons that led to the creation of the GIP-PDS, which is based in particular on the health data warehouse of the French National Health Data System (SNDS) (see appendices 4.1 and 4.2), illustrate the challenges involved in creating HDPs, some of which already existed.

3.1 Reasons for creating health data platforms

3.1.1 Towards the 4 Ps of medicine

The field of medicine is undergoing a series of revolutions that should enable it to respond more effectively to the health challenges facing our societies. The first revolution concerns precision medicine, which involves developing targeted interventions that take account of a person's genetic and environmental profile. The second revolution is the possibility of better identifying risk, protection and resilience factors, which will make it possible to prevent, predict and diagnose the emergence of disease in populations at greater risk of developing specific illnesses. Finally, the third revolution is the development of

[28] IoT: Internet of Things.

[29] See: https://iqvia.opendatasoft.com/pages/accueil/ et https://www.data.gouv.fr/fr/organizations/iqvia-france/.

[30] See: https://www.cegedim.fr/.

[31] See: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868. Published in the Official Journal of the European Union on 23 June 2022, the DGA will come into force in September 2023.

participatory medicine, co-developed with all stakeholders (patients, carers, medical and administrative staff). These developments towards what is known as the 4 Ps of medicine - personalised, preventive, predictive and participatory - require the use of large-scale, high-quality data.

3.1.2 Issues relating to research, the public interest and private organisation

The GIP-PDS was created in November 2019 to facilitate the sharing and use of health data from a wide variety of sources in order to promote research in this field. Its creation followed the recommendations of the Villani[32] report on artificial intelligence (AI), which advocated the sharing of health data, and meets the desire of the President of the Republic to make France an international leader in the field of digital health.

Generally speaking, HDPs bring together health data from a wide variety of sources, which are generally dispersed and heterogeneous, to facilitate their joint use. Bringing this data together on a platform makes it accessible and standardised, and enables it to be processed digitally using computing power or sophisticated AI algorithms such as deep learning.

The cross-referencing and processing of a wide variety of health data can be used to meet a number of objectives: to advance biomedical research and innovation, medical decision-making (prescriptions, interpretation of biological tests), clinical care, health monitoring and medical device vigilance (monitoring of medical devices under real-life conditions), and finally to help manage the healthcare system.

3.2 Protection of the individual, public interest and the common good

3.2.1 Scientific research and the public interest

The use of health data in the public interest must ensure the protection of personal data. The GDPR advocates individual consent, with the exceptions mentioned above (section I.1.2). However, in the case of scientific research, the scope of some of these definitions needs to be clarified. As Shabani points out[33], the GDPR gives little guidance as to what can be considered scientific research of public and general interest, particularly where commercial entities are involved. While the GDPR considers research supported by private funds to be scientific research, it does not distinguish whether it is profit-driven or not, nor does it indicate whether the public interest takes precedence over private and commercial interests. Furthermore, the interpretation of the research exemption in the countries covered by the GDPR reveals a wide disparity[34]: 18 countries, including France, have developed specific regulations on research and the public interest, while only nine countries have adopted specific provisions in the case of research conducted by private bodies.

In France, Article 66 of the Data Protection Act[35] states that "The guarantee of high standards of quality and safety of healthcare and medicines or medical devices constitutes

---

[32] Villani C., (2018), *Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne*, 235 p. [https://www.vie-publique.fr/sites/default/files/rapport/pdf/184000159.pdf].

[33] Shabani M., (2022), "Will the European Health Data Space change data sharing rules?", *Science*, vol. 375, Issue 6587.

[34] European commission, Assessment of the EU Member States' rules on health data in the light of GDPR, (2021):[https://health.ec.europa.eu/system/files/2021-02/ms_rules_health-data_en_0.pdf]

[35] Law no. 78-17 of 6 January 1978 on data processing, data files and individual liberties [https://www.legifrance.gouv.fr/loda/article_lc/LEGIARTI000038888793/].

a public interest purpose". The CNIL specifies what is meant by a public-interest mission[36]: "The public-interest mission is one of the six legal bases provided for by the GDPR authorising the implementation of personal data processing. [...] This legal basis therefore primarily concerns processing carried out by public authorities. It may, however, authorise the implementation of processing by private bodies, provided that they pursue a public-interest mission or are endowed with prerogatives of public authority", such as private health establishments entrusted with a public service mission. This legal basis means that consent is not required.

In the health sector, the GIP-PDS provides guidance to project leaders - who may be health industry operators or insurers - to help them identify whether their project pursues a public interest objective, by referring to the SNDS[37]:

> "In addition to pursuing a public interest purpose, projects must verify the purposes specific to the SNDS, i.e. to contribute (i) to information on health, as well as on the supply of care, the provision of medico-social care and its quality; (ii) to the definition, implementation and evaluation of health and social protection policies; (iii) to understanding health expenditure, health insurance expenditure and medico-social expenditure; (iv) to informing health and medico-social professionals, structures and establishments about their activities; (v) to health surveillance, monitoring and safety; (vi) to research, studies, evaluation and innovation in the fields of health and medico-social care".

In addition, the GIP-PDS specifies that it is forbidden to carry out processing "either for the purpose of taking a decision against an individual identified on the basis of data concerning him/her and appearing in one of these processing operations, or for the purpose of promoting health products to health professionals or establishments, or excluding cover from insurance contracts or modifying insurance contributions or premiums for an individual or group of individuals".

3.2.2 Commodification of personal data or common good

Although the entry into force of the GDPR put an end to the debate on the legal status of personal data in Europe, by considerably improving its protection, the fact remains that there is also a market in Europe for the exchange of data that is sometimes called "medical" to distinguish it from "health" data. The data traded must have been anonymised (see section I.4.6). A data broker (see section I.2.5) may therefore legally sell, for example, a thousand X-rays of a fractured wrist, provided that they do not contain any information that could be used to re-identify the individual.

On the other hand, the right of ownership and transfer of health data (which are therefore identifying) is practised in some countries by the same data brokers. For the advocates of this development, it is not so much the principle of transfer that is at the heart of the debate, but that of transfer against payment. Recognising a right of transfer is not in itself essential, since personal data can be communicated to third parties. But granting a right of ownership over personal data would make it possible to recognise its financial value. Its transfer could then be subject to remuneration.

Many institutions have come out against such a change in the law, pointing out the scale of the task (defining a new legal regime) for little financial gain. Data from a single individual, unlike aggregated or enriched data, has little market value. More generally, from an ethical point of view, we may question the sale and therefore the commercialisation of

---

[36] CNIL, "*La mission d'intérêt public : dans quels cas fonder un traitement sur cette base légale ?*" » [https://www.cnil.fr/fr/les-bases-legales/mission-interet-public].

[37] Health Data Hub, "*Qu'est-ce que l'intérêt public ?*" [https://health-data-hub.fr/interet-public].

information that is hitherto considered to be intimately linked to the individual and his or her identity.

There remains the little-explored and radically contrasting path of the commons, inspired by the work on institutional economics of Elinor Ostrom (winner of the 2009 Nobel Prize in Economics). Commons are shared resources the management of which is based on the establishment of a system of socially sanctioned rules by their users. Free software is the best example of digital commons. Elinor Ostrom has developed a theory of the knowledge commons that applies to the management of intangible resources such as information. Translating this economic concept into a legal concept, which is unprecedented in France as such (with the exception of communal "sectional property"), would require the creation of a legal instrument that could be inspired by similar concepts, in particular the *choses communes* (commons) governed by Article 714 of the French Civil Code. However, such a development seems better suited to the legal status of aggregated data or data networks than of the data itself. This development would be intellectually appealing, but the practical differences it would bring to data management are not easy to envisage.

3.3 Examples of projects supported by health data platforms

To illustrate the value of bringing together a variety of health data on HDPs, we shall present a few examples of projects which use either the GIP-PDS presented in the appendix (Appendix 4.2), or large cohorts such as Constances (Appendix 4.9) or UK Biobank (Appendix 4.10), which have no specific characteristic other than having been easily identified and presented by the members of this working group. It should be noted that as far as the GIP-PDS is concerned, there are as yet no completed projects whose results could be evaluated. For each project, the data is processed in a space specific to the project, which contains only the necessary data. The project manager has remote access to the data and cannot retrieve it.

3.3.1 HYDRO (GIP-PDS)

The HYDRO[38] project (Development and validation of algorithms for predicting heart failure attacks in patients with connected implants) is run by the private company Implicity[39]. It received all necessary authorisations and began in October 2021 for a period of five years. It matches data from the SNDS with data from the Implicity platform used by cardiologists for remote patient monitoring, according to a process[40] validated by the GIP-PDS.

3.3.2 Glucocorticoids (GIP-PDS)

The Glucocorticoids[41] project is interesting, even if it does not involve data matching, as it is an international project initiated by the European Medicines Agency (EMA) and involving seven countries. The data controller is the American multinational IQVIA, a specialist in health data. The aim of the study is to improve the management of patients with SARS-CoV-2 or suspected SARS-CoV-2. The aim is to use public health data from the SNDS to determine which patients could benefit most from the use of glucocorticoids, and to choose

[38] Health Data Hub, "*HYDRO : Développement et validation d'algorithmes de prédiction des crises d'Insuffisance cardiaque chez les patients porteurs d'implants connectés*" [https://www.health-data-hub.fr/projets/hydro-developpement-et-validation-dalgorithmes-de-prediction-des-crises-dinsuffisance].

[39] See the website: https://www.implicity.com/.

[40] See https://www.health-data-hub.fr/sites/default/files/2021-10/Infographies%20Projet%20HYDRO.pdf

[41] Health Data Hub, "*Utilisation des glucocorticoïdes par voie systémique dans le traitement de la COVID-19 et risques d'événements indésirables*" [https://health-data-hub.fr/projets/utilisation-des-glucocorticoides-par-voie-systemique-dans-le-traitement-de-la-covid-19-et].

the best molecules, dosage and time of administration. The aim of the study is to describe patterns of systemic glucocorticoid use and adverse events associated with these drugs in a cohort of SARS-CoV-2 patients during 2019 and 2020. It also plans to validate the feasibility of European studies on SARS-CoV-2 treatments using data put into the international OMOP[42] format by HDP engineers, at the request of the EMA.

3.3.3 REXETRIS (GIP-PDS)

The REXETRIS[43] project (*Relations EXposition - Effet à long terme chez le Transplanté Rénal des médicaments ImmunoSuppresseurs*) is led by Limoges University Hospital and supported by Optim'Care. Its aim is to improve monitoring of kidney transplant patients by proposing new ways of optimising more targeted immunosuppressive treatments. To achieve this, it is studying a retrospective cohort using data from three databases: CRISTAL from the Biomedicine Agency (ABM), SNDS (CEPIDC database) from the CNAMTS and ABIS from the Limoges University Hospital, a database used in patient management. The project envisages two matching operations to create a pseudonymised database of all kidney transplant patients followed in a French transplant centre since 2005: between ABIS and CRISTAL, and between CRISTAL and SNDS. It received all the necessary authorisations for processing and began in July 2021 for a period of five years. Data will be kept for seven years once the project has been completed.

3.3.4 Chronic diseases (Constances)

The Constances HDP, based on a large French cohort, has enabled a number of research projects to be carried out on chronic diseases, including sexual activity in diabetic women, chronic obstructive pulmonary disease and HIV infection, and alcohol-related morbidity.[44]

3.3.5 Neuroanatomy of the brain (UK Biobank)

The UK Biobank HDP, based on a large British cohort, has produced significant results on brain neuroanatomy by linking epidemiological and genetic data with magnetic resonance imaging (MRI) data of the brain, heart and abdomen from 100,000 participants. The combined analysis of genotyping and brain MRI data made it possible, among other things, to estimate the heritability of inter-individual differences[45, 46].

3.4 Three initial ethical issues linked to mass data collection

The volume of data collected in HDPs must be sufficiently large and of sufficient quality to enable meaningful processing, such as that carried out using machine learning, whether it be data for primary use, collected from patients, such as that in the SNDS or *Mon espace santé*, or health data for secondary use from research projects, the quality of which depends on the data for primary use, the processing algorithms and the 'curation' operations they have undergone (i.e. maintenance and cleaning work). However, it is worth questioning the wisdom of seeking to store data on a mass scale in order to build up the most comprehensive databases possible, in accordance with what might be called a

---

42 See: https://www.ohdsi.org/data-standardization/the-common-data-model/.

43 Health Data Hub, "*REXETRIS : Relations EXposition - Effet à long terme chez le Transplanté Rénal des médicaments ImmunoSuppresseurs*" [https://www.health-data-hub.fr/projets/rexetris-relations-exposition-effet-long-terme-chez-le-transplante-renal-des-medicaments].

44 See: https://www.constances.fr/projets-termines.

45 Biton, A. *et al.*, (2020), "Polygenic Architecture of Human Neuroanatomical Diversity", *Cereb. Cortex N. Y. N 1991,* **30**, 2307–2320.

46 Elliott, L. T. *et al.*, (2018), "Genome-wide association studies of brain imaging phenotypes in UK Biobank", *Nature* **562**, 210–216.

reverse precautionary principle. The advent of Big Data has given rise to an overabundance of data, which raises three ethical tensions that call into question the epistemic model of Big Data in healthcare[47].

The effectiveness and relevance of treatments carried out on HDPs depend not only on the volume but also on the quality of the data used. As the HTF-Sopra Steria Next report points out, one of the principal determinants of data quality is its representativeness[48]. According to the Artificial Intelligence & Cancers Association[49], the more varied the sources of data, the lower the risk of bias in cancer research. Quality is intrinsically linked to data diversity.

But HDP users can only find what is stored in the HDPs (unless they request new data for their study, after obtaining consent from the individuals concerned). There is therefore a tension between the representativeness of the data (to avoid any bias), and respect for the autonomy of individuals, who may or may not give their consent (see sections III.1 to III.4) for their data to be used and stored. This means that the diversity and representativeness of the data cannot always be guaranteed. Therefore, in order to make informed use of HDP data, HDP designers must specify precisely what is stored in the HDP (see Recommendation 1).

HDP users are also responsible for choosing their dataset carefully, taking into account the purposes of their project, and for remedying any bias, as, for example, the Dawex data exchange platform[50] does, offering sampling tools to automatically generate representative data examples based on algorithms, in order to avoid any bias[51]. If it proves impossible to avoid biases, a weighting method to take them into account can be implemented, as UK BioBank[52] does, for example (see Recommendation 1).

The need to store mass data to use effective digital tools on HDPs such as machine learning systems conflicts with the principle of minimising collection, storage and retention period, as prescribed by the GDPR. In addition, the volume of data susceptible to hacking due to security flaws or malicious use increases with the volume and duration of data storage. This poses a risk to the protection of sensitive personal data and respect for privacy.

It is therefore impossible to overstate the need to respect the principle of proportionality when collecting data for specific purposes (see the GDPR's minimisation principle), including when collecting data for research purposes. Nevertheless, minimising the retention period for health data collected in the event of a health crisis, which is a matter for the legislator to decide, must take account of the needs of research (see Recommendation 2).

Finally, even if the impact of digital technology on the environment, and in particular the impact of cloud services, seems difficult to assess, maximising the storage of health data in order to improve therapeutic results for the common good runs counter to digital sobriety, which aims to limit digital storage and calculations in order to limit the impact on

---

47 In this opinion, we do not address the ethical issues raised by the increasing digitisation of health data and tools. We limit ourselves here to the ethical issues raised by the storage and exchange of mass health data via health data platforms.

48 HTF–Sopra Steria Next report, (2022), "*Data-altruisme, une initiative européenne. Les données au service de l'intérêt général*", Human Technology Foundation and Exploratoire Sopra Steria Next report, p 26.

49 See: https://filiere-ia.fr/.

50 See: https://www.dawex.com/ and Appendix 4.11

51 See: https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained#ecl-inpage-l4ihmeih

52 See: https://www.medrxiv.org/content/10.1101/2022.05.16.22275048v1.full

the environment. We therefore need to assess the environmental impact of current and future HDPs and seek to minimise it, particularly when they are used for research, as recommended in a recent opinion by the CNRS Ethics Committee[53] (see Recommendation 3).

4. Fundamental principles for building architectures

In this section, we focus on the ethical issues associated with the IT solutions chosen for the design of HDPs. Ethical issues relating to modes of governance and business models are addressed in the following sections.

As soon as the architecture of an HDP is designed, three fundamental technical principles with ethical implications must be respected: security, interoperability and portability. These three principles are at the heart of the GDPR.

4.1 Security

The **security** of infrastructures and protocols is a crucial technical requirement for data platforms in general, and HDPs in particular, due to the sensitive nature of health data. It concerns:

- HDP storage conditions, to ensure data integrity (no unauthorised person can modify it) and confidentiality (no one can extract it without authorisation);
- data transfer conditions, whether this involves uploading data to the HDP from disparate sources or using the data to perform off-site calculations (download) or exchange data with other HDPs.

The potential for breaches of HDP content, as a result of security breaches or malicious use, raises three ethical issues: confidentiality, non-malevolence and sovereignty.

It is worth recalling the existence of a public solution, the CASD (Centre d'accès sécurisé aux données - Secure Data Access Centre)[54], which hosts highly sensitive data within the meaning of the GDPR from public operators and is a secure data access hub. It can therefore be used for biomedical data. The *Constances* cohort (see Appendix 4.9) developed by Inserm uses the CASD for its research projects via secure bubbles.

4.2 Interoperability

Interoperability refers to the possibility of exchanging health data between different HDPs. It encourages exchanges, but has to contend with security issues intrinsic to the exchange of data between HDPs that do not rely on the same infrastructures and do not use the same software. According to an analysis by the Gaia-X association, lack of interoperability, portability (see section I.3.4) and data sovereignty (see section II.1) are the main reasons that seem to have prevented faster adoption of cloud computing in Europe[55].

---

[53] COMETS – CNRS Ethics Committee, "*Intégrer les enjeux environnementaux à la conduite de la recherche – Une responsabilité éthique*", Opinion no. 2022-43, 5 December 2022. See: https://comite-ethique.cnrs.fr/avis-du-comets-integrer-les-enjeux-environnementaux-a-la-conduite-de-la-recherche-une-responsabilite-ethique/

[54] See: https://www.casd.eu/ and Appendix 4.5.

[55] Tardieu H. et al., (2022), "Compliance, and consequences on the labeling framework **of** Gaia-X – See: https://gaia-x.eu/wp-content/uploads/2022/07/Gaia-X-Compliance-Document_Final_f.pdf

Interoperability presupposes an effort to **standardise** health information systems by implementing standards for IT systems, data processing software and data representation formats.[56]

With regard to the *Ouest Data Hub* (ODH), interoperability by design has been targeted so as to enable effective communication with the six HDWs in the HUGO network, based on a common technology.

Similarly, the French Digital Health Agency (ANS), with a view to the creation of *Mon espace santé* for all and its use by all doctors, has defined interoperability standards to facilitate the digitisation and standardisation of information exchanged, and has begun by modifying data formatting and storage software to anticipate data exchange.

Interoperability also involves the development of healthcare terminologies. With regard to oncology data, the interSIRIC OSIRIS consortium initiative to federate databases specialising in this field has resulted in a dynamic standardised model for representing oncology data[57].

Also worth mentioning is the Oncolab[58] research-innovation project, launched on 1 July 2022 by a French public-private consortium, bringing together the companies Arkhn and Owkin, the INRIA research institute and hospitals specialising in cancer treatment: IUCT-Oncopole, Instituts Curie in Paris and Bergonié in Bordeaux, as well as the Toulouse University Hospital. The aim of the project is to standardise access to health data for oncology research.

The aim of the OMOP-CDM (Observational medical outcomes partnership - Common Data Model)[59] is to ensure **interoperability** between the various health databases, whether clinical or medico-administrative. It is the result of a public-private partnership called OMOP (Observational Medical Outcomes Partnership), chaired by the US Food and Drug Administration (FDA) and funded by a consortium of pharmaceutical companies set up in 2008 for a five-year period. It is used by the GIP-PDS.

The FHIR (Fast Healthcare Interoperability Resources)[60] communication standard is a standard describing data formats and other elements as well as an application programming interface for the exchange of medical information. The standard was developed by Health Level Seven International (HL7), a not-for-profit organisation dedicated to the development of health data interoperability and the standardisation of medical exchange protocols.

With regard to genomic data, the Global Alliance for Genomics and Health (GA4GH[61]) proposes various open standards for sharing biomedical data.[62] [63]

All these observations only serve to reinforce the importance of HDP interoperability, and it would be desirable for France to play a greater role in the European effort to develop

---

[56] See, for example, the Inserm dossier: "*Big data en santé : des défis techniques et éthiques à relever*" published 27/06/2022 [https://www.inserm.fr/dossier/big-data-en-sante/].

[57] See: https://siric-brio.com/premiers-resultats-du-consortium-intersiric-osiris/.

[58] See: https://www.bioworld.com/articles/520331-consortium-launches-oncolab-to-standardize-access-to-oncology-data?v=preview.

[59] See: https://www.ohdsi.org/data-standardization/the-common-data-model/.

[60] See: https://fr.wikipedia.org/wiki/Fast_Healthcare_Interoperability_Resources

[61] See: https://www.ga4gh.org/.

[62] See: https://www.ga4gh.org/genomic-data-toolkit/.

[63] Rehm H.L. et al., (2021), GA4GH: International policies and standards for data sharing across genomic research and healthcare", *Cell Genom*, 10; 1(2): 100029.

standards and norms for formatting and structuring health data[64] (see Recommendation 4).

4.3 Reversibility and portability

Reversibility is the possibility of changing service provider (supplier of a hosting service) or re-internalising databases and associated processing, with minimal cost in terms of applications, data and infrastructure. The GIP-PDS has been working on reversibility since its pre-configuration in 2019. Although it has chosen the Microsoft Azure cloud solution, which is the only one available in the short term with HDS (Health Data Hosting) certification and which can integrate the functionalities and certifications necessary for the required level of security, the objective of reversibility has been part of its roadmap since 2022[65].However, reversibility can only be achieved through the **portability** of data and IT programmes.

Data portability is defined in the GDPR[66]: "The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided, where: the processing is based on consent (...), or on a contract (...) and the processing is carried out by automated means."[67] The CNIL specifies that portable data, collected with the consent of the data subject or under contract, must be provided "in a structured, commonly used and machine-readable format". This means that the organisation must offer data formats adapted to the type of data concerned, giving preference to open, interoperable formats" [68].

With regard to the portability of a computer program, this is its capacity to be adapted more or less easily in order to function in different execution environments. The differences may relate to the hardware environment (processor) or the software environment (operating system). The difference in environment may also relate to a combination of both elements. This is the case, for example, in the fields of embedded computing, supercomputers and virtual machines.

The portability of algorithms and programmes used on platforms processing health data is an important ethical issue in terms of the reproducibility of calculations and the ability to benefit from the skills and efforts developed on a platform for the benefit of as many people as possible. It also touches on issues of transparency and explicability, and can benefit greatly from open source approaches (see section I.4.5). Portability is a complex and long-standing issue[69], but one that is taking on new importance in the current context of digital sovereignty. It depends on the hardware and software architectures specific to each platform, in a context of complex visualisation algorithms, machine learning, database management and analysis, the use of computer networks and cyber-security. This portability is still a subject of scientific and technological research, as well as a standardisation issue. It must be taken into account from the design of algorithms through

---

[64] See: https://www.snomed.org/news-and-events/articles/EU-drives-standardized-terminology-funding-program.

[65] See: https://www.health-data-hub.fr/sites/default/files/2022-01/HDH_Feuille_De_Route_2022_0.pdf.

[66] Article 20 of Chapter 3 of the GDPR: https://rgpd.com/gdpr/chapter-3-rights-of-the-data-subject/article-20-right-to-data-portability/

[67] The conditions are: "where processing is based on consent pursuant to Article 6(1)(a) or Article 9(2)(a), or on a contract pursuant to Article 6(1)(b)".

[68] See: https://www.cnil.fr/fr/le-droit-la-portabilite-obtenir-et-reutiliser-une-copie-de-vos-donnees

[69] See https://www.cigref.fr/archives/histoire-cigref/wp-content/uploads/2017/02/CIGREF-1977-portabilite-applications-informatiques.pdf.

to their programming on different hardware architectures. This last point is tricky because the machines used implement architectures that can be very different from one another. In particular, we need to distinguish between situations where programmes are used on standard machines (typically individual workstations) and those where supercomputers are used.

The issue of portability of data, algorithms and programmes is therefore strategic for the reversibility of HDP contracts, as it is for interoperability (see Recommendation 4).

4.4 Centralised or non-centralised architecture

While all existing HDPs generally aim to comply with the four principles discussed above, they differ - among other things - in the way in which they have been built up from existing health databases, which has ethical implications in particular.

The GIP-PDS[70] has opted for a centralised platform. It was set up in 2019 on the basis of the National Health Data System (SNDS) and by copying the health databases listed in its catalogue into its centralised, secure platform. The GIP-PDS can also receive data from sources not listed in the catalogue. The data is stored in the Microsoft Azure[71] cloud, which has given rise to considerable controversy regarding respect for data sovereignty (see section II.1.3).

The *Ouest Data Hub* (ODH)[72] is an inter-regional hub also created in 2019, but which from the outset adopted a more decentralised and collaborative vision, choosing to network the HDWs of several hospital centres, and drawing on their local expertise in clinical data and their clinicians. These HDWs have been developed in the most homogenous way possible, using a common technology developed within Inserm, in conjunction with the Rennes University Hospital to integrate the data, based on a public-private partnership with the company Enovacom. Only the data required for the projects is stored on the HDP infrastructure, which is hosted by the Nantes University Hospital and recognised as a Health Data Host (HDS).

It is currently not possible to fully assess the advantages and disadvantages of centralised approaches (with storage in a cloud) compared with non-centralised approaches (inter-regional hub with HDP infrastructure hosted locally in a university hospital) in terms of data **security** and **confidentiality**.

Storing data in a single location means that only one safe needs to be monitored, but if this safe is breached, the loss of data is greater than in a decentralised approach, where an attack on one of several safes would affect less data. Consideration should be given to simulating security breaches in order to better assess the risks of each of these approaches (see Recommendation 5).

A decentralised approach such as that of the ODH makes it possible to draw on existing IT systems (the six HDWs of the university hospitals in the Western region). This requires the use of a common technology for the development of local HDWs, as the ODH has done, but it also makes it possible to take into account and **respect** the work being done on the IT systems of local hospitals and to allow them a degree of **autonomy** in the management of their warehouses. In addition, the creation of the ODH was accompanied by the launch of

[70] See Appendix 4.2 and https://www.health-data-hub.fr/.

[71] The French National Agency for Information Systems Security (ANSSI) has developed a set of requirements for SecNumCloud cloud computing service providers, and has provided a list of service providers qualified under these requirements. The approved service providers include two French companies: OVH and Outscale SAS.

[72] See Appendix 4.3 and https://www.chu-hugo.fr/accueil/projets/Ouest-DataHub/

multicentre projects, highlighting the need to share data between hospitals, which created a virtuous circle and encouraged clinicians to join the ODH project (see Recommendation 6).

As far as the use of HDPs is concerned, a centralised approach that makes all the data required for algorithmic processing available in one place seems to be more favourable to the exploitation of data by machine learning algorithms. Advocates of a decentralised or federated approach are banking on advances in research and innovation in the field of federated artificial intelligence (see Recommendation 5).

In both types of approach, the cost of the associated infrastructure, research and innovation, as well as the need for human resources, are considerable and must be supported.

4.5 Promoting open technical solutions

As the French National Authority for Health (HAS) points out, "the issue of private ownership of data schemas, particularly in the case of electronic patient record managers, is a significant concern" [73], not to mention the fact that conflicts of interest and ethical problems are likely to arise. Indeed, these private operators often favour the use of proprietary licences and closed source code, which also poses a number of problems, in particular a lack of transparency and a strong dependence on the publisher for maintenance and adaptation to new products.

This is why we believe it is essential to promote the development of open solutions and open source technologies for HDPs as in other areas. They offer code transparency by default and allow all players to use them legally. Their reputation is occasionally tarnished by claims that they are less reliable, but numerous experiences have shown the opposite to be true. Furthermore, open source is an important factor in attracting talent that is often drawn to this model[74]. It is for this reason that we endorse the HAS recommendation to encourage the emergence of open technology platforms[75] (see Recommendation 7).

4.6 Pseudonymisation and anonymisation of data

Most HDPs pseudonymise the health data they collect. In doing so, this sensitive data remains personal data which is covered by the GDPR. Anonymising data would be a better guarantee of patient privacy. However, the possibilities for cross-referencing health data with other databases are increasing, and re-identification is becoming more and more feasible, particularly in the case of rare diseases, which calls into question the very possibility of anonymisation. Furthermore, in a certain number of cases, it can be of benefit to find the patients whose data is being analysed. There is therefore a tension between anonymising data to respect patient privacy and preserving their identity for better care.

In the case of HDPs collecting genomic data, the situation is even more delicate because the data is highly identifiable and concerns not just the individual but the whole family, while at the same time being fundamental to the development of precision medicine. On

---

[73] See: https://www.has-sante.fr/jcms/p_3386076/fr/entrepots-de-donnees-de-sante-hospitaliers-la-has-publie-un-panorama-inedit-en-france, p. 17. https://www.has-sante.fr/upload/docs/application/pdf/2022-11/rapport_entrepots_donnes_sante_hospitaliers.pdf page 28 ??

[74] Alcaras, Gabriel. "*Des logiciels libres au contrôle du code. L'industrialisation de l'écriture informatique*". Doctoral thesis, Paris, EHESS, 2022. https://www.theses.fr/s341603.

[75] See: https://www.has-sante.fr/upload/docs/application/pdf/2022-11/rapport_entrepots_donnes_sante_hospitaliers.pdf, page 30.

this subject, Bonomi and his co-authors[76] take stock of the main threats to privacy in the collection and use of genomic data and the protection techniques that exist or need to be developed.

The HTF-Sopra Steria Next report[77] refers to the unsuccessful initiative of the Robert Koch Institute, which set up the Corona Data Donation application in Germany in April 2020, in the early months of the pandemic. With this application, people were invited to share their health data, in particular symptoms linked to contamination by the virus, and the pseudonymisation of the data was guaranteed. A year later, the German Federal Commissioner for the Protection and Freedom of Information noted that only one million users had adopted this application, attributing this low take-up to the ambiguity of the notion of "data donation and the choice of pseudonymisation over total anonymity". This suggests that anonymisation methods need to be developed to instil confidence in patients who are likely to entrust their health data to an HDP. Alongside research into new anonymisation techniques, other avenues are being investigated, such as the exploration of homomorphic techniques[78] or the design of algorithms for generating anonymous synthetic data - avatars - from personal data[79] (see Recommendation 8). However, no anonymisation procedure has yet been certified, particularly by the CNIL.

4.7 Best practice

Recent research[80] has reviewed the approaches taken by scientists to sharing the data they use. It shows that, generally speaking, scientists adopt approaches that comply with ethical and legal principles, but that they bemoan a lack of practical procedures that would, by default, guarantee the ethical and legal collection and sharing of data. A set of technological solutions enabling ethics by design for health data platforms is called for.

5. Recommendations

Data quality and sharing

- **Recommendation 1:** Explain the nature and origin of personal health data collected in HDPs, distinguishing between their primary and secondary uses and, for a given research project, use unbiased datasets, or, where this is not possible, take account of these biases in their analysis, for example, through weighting methods.
- **Recommendation 2:** Ensure that the retention period for the public health data collected is properly calibrated in relation to the requirements of the research, without neglecting the necessary protection of personal data.

Environmental impact of HDPs

- **Recommendation 3**: Evaluate the environmental impact of HDPs and aim for energy sobriety through appropriate choices of data storage, architecture and operating modes.

---

[76] Bonomi L., Huang Y., & Ohno-Machado L., (2020), "Privacy Challenges and Research Opportunities for Genomic Data Sharing", Nat. Genet 52, 646–654.

[77] HTF-Sopra Steria Next report, (2022), "*Data-altruisme, une initiative européenne. Les données au service de l'intérêt général*", Human Technology Foundation and Exploratoire Sopra Steria Next report, p 28.

[78] Gentry C., (2009), "A fully homomorphic encryption scheme", PhD thesis, Stanford University.

[79] See, for example: https://www.larevuedudigital.com/anonymisation-des-donnees-de-sante-experimentee-au-chu-de-brest-avec-une-startup/.

[80] Johansson V., et al., (2022), "What ethical approaches are used by scientists when sharing health data? An interview study", *BMC Medical Ethics*, 23:41.

HDP architecture:

- Recommendation 4: Ask public authorities to become more involved in the development of standards and norms for formatting and structuring health data in order to promote better portability and interoperability of HDPs.
- Recommendation 5: Carry out comparative evaluation studies between centralised and decentralised approaches to HDPs and their combinations, to ensure secure management of health data. Encourage innovations in federated Artificial Intelligence to inform the debate between centralised and decentralised architectures.
- Recommendation 6: Choose HDP architecture solutions that respect local ecosystems and take into account multi-centre research projects that require data distributed across various clinical centres or hospitals, highlighting the benefits of pooling data.
- Recommendation 7: Encourage public HDP creators to adopt open standard formats and open source algorithms to enhance data quality and subsequently process data flows, and also enable multi-centre studies, in order to release the innovation potential of all reusers of health data.

Anonymisation:

- Recommendation 8: Develop research into alternative methods to data anonymisation and pseudonymisation, in particular homomorphic encryption techniques, in order to make better use of health data.

II. SOVEREIGNTY, AUTONOMY AND VALUATION OF HEALTH DATA PLATFORMS

1. Ethical issues surrounding the sovereignty of health data platforms

In an article published in La Croix on 28 June 2022[81], journalist Marion Durand refers to the results of an Ifop[82] survey according to which 52% of French people do not trust any country to protect their personal data, and only 10% would prefer a European player. She argues that the suspicion of the French people has been fuelled by the various uses of personal data for commercial or political purposes and recent cases of data leaks (up by 19% in 2020 according to the International Cybersecurity Forum), particularly of medical data.

1.1 The complex geopolitics of health data

The tensions we observe at national level, between the need and the desire to protect individual data while making it accessible for the development of science and projects of common interest, are mirrored at international level. Each country has an interest in protecting and therefore limiting access to its own data, while at the same time benefiting from the data of others and therefore sharing its own data in order to contribute to the advancement of knowledge. These tensions take specific forms at the geopolitical level. Firstly, there is a difference in legal model between the European GDPR, which sees health data as a personal attribute, and the more liberal US model, which sees it as a commodity that can be marketed. This can lead to complex situations, such as that of Ireland, which is seeking to attract platforms, including American ones, to its territory and is therefore finding it difficult to apply European privacy regulations[83]. Conversely, it is important for a country's players to be able to join international research and development networks, including with their data, in order to participate in the global advances they can generate[84].

Faced with these difficulties, public debate in France has focused on the notion of defending data sovereignty.

1.2 Ambivalence of the notion of sovereignty

Sovereignty is a complex concept that stems from law, philosophy and political science. The classical, 'closed' concept of sovereignty refers to power exercised over a territory protected by borders. Since the modern era, it has been associated with and confronted by the entrepreneurial and 'open' concept of deterritorialised sovereignty of economic actors controlling financial and commercial flows.

While the classical concept is associated with an idea of protection that could be unilaterally imposed, the entrepreneurial concept requires the defence of sovereignty to be negotiated, often within a multilateral international framework.

---

81 Durand M., "*L'altruisme des données, une utopie ?*" , *La Croix*, 28 June 2022.

82 Ifop poll, *Les Français et la souveraineté numérique*, April 2021.

83 See: https://www.lemonde.fr/pixels/article/2021/09/13/protection-des-donnees-l-irlande-maillon-faible-du-rgpd_6094434_4408996.html

84 See: https://www.udninternational.org/ Undiagnosed Diseases Network International is an example of an international genomic data network aimed at improving the diagnosis of very rare diseases. Geneticists from all countries are welcome to join, provided they share clinical cases.

This distinction and ambivalence are reflected in the notion of digital sovereignty[85]. It is interesting to note that the Villani report on AI[86] never mentions "digital sovereignty", but encompasses it within the broader issue of "technological and economic sovereignty".

This ambivalence can also be found when it comes to health data, which is seen as both a "national treasure" to be protected and a "common good" to be shared, on a European or even global scale. The centralised Social Security system introduced by the law of 26 January 2016 has had the unexpected beneficial effect of facilitating the construction of a virtually exhaustive database on the health of the French population. This National Health Data System (SNDS[87]) now represents one of the most comprehensive and richest sources of data in the world. But the question of how best to exploit this source of information means rethinking the role of the State in the conduct of innovation policies, and in particular the defence of its sovereignty insofar as healthcare is a fundamental mission of the French State, which led to the GIP-PDS project[88]. These issues of sovereignty over health data can be found on the smaller scales of a region, with the example of the *Ouest Data Hub*[89] (Appendix 4.3), or university hospitals (CHU), which want to both protect and exploit their own health data. The same concerns are shared by the European Union, which, through the Gaia-X European Association for Data and Cloud[90], aims to promote the values of data protection, transparency, security and respect for data rights.

1.3 A liberal and entrepreneurial vision for conquering sovereignty

The consideration of sovereignty in terms of data exploitation has led to the promotion of a liberal and entrepreneurial vision of health data platforms (HDP). The impetus given by the Villani report on Artificial Intelligence sees health data first and foremost as enabling France to become a "world leader in digital health" in two complementary ways. Firstly, this abundance of available data should make France a very attractive location worldwide, where existing French and foreign companies developing digital health services and researchers working in this area could find data to work with. Secondly, such a health data ecosystem should generate breakthrough innovations in France on the scientific side and, on the business side, 'unicorns', i.e. start-ups valued at over 1 billion dollars, which should enable them to compete with the so-called GAFAM[91], or at the very least with the Anglo-Saxon multinationals. These unicorns would thus constitute the main weapon - or rather the main defensive horn - of France and Europe in international economic competition. Sovereignty, here, means the national capacity to resist the domination of foreign (mainly American) economic players by developing our own economic power on an international scale.

In France, this liberal vision accommodates the central role played by the State in defending this sovereignty, as reflected, for example, in MP Eric Bothorel's report entitled

---

85 Ganascia J-G., Germain E., Kirchner C., (2018), *La souveraineté à l'ère du numérique Rester maîtres de nos choix et de nos valeurs*, CERNA, 36 p. http://cerna-ethics-allistene.org/digitalAssets/55/55708_AvisSouverainete-CERNA-2018.pdf

86 Villani C., (2018), *Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne*, 235 p. [https://www.vie-publique.fr/sites/default/files/rapport/pdf/184000159.pdf].

87 See Appendix 4.1 and https://www.snds.gouv.fr/.

88 See Appendix 4.2 and https://www.health-data-hub.fr/

89 See: https://www.chu-hugo.fr/accueil/projets/Ouest-DataHub/

90 See: https://www.gaia-x.eu/

91 GAFAM is the acronym for the five major American web companies: Google, Apple, Facebook, Amazon and Microsoft.

Pour une nouvelle ère de la politique publique de la donnée[92] [For a new era in public data policy]. As Mariana Mazzucato[93] has written, the State should make the initial investments that companies avoid because they are too risky, in order to develop not just one company, but an entire market. To achieve this, the government has, on the one hand, set up the GIP-PDS, endowed with almost €10 million in 2020, whose project is to centralise health data and make it easily accessible to researchers and start-ups and, on the other hand, established, through the Banque Public d'Investissement (Bpifrance), a financing plan called DeepTech, whose most structured sector, health, has injected more than €100 million since 2019 into the financing of start-ups, some of which rely precisely on the data offered by the GIP-PDS. In this way, the State has built a "sandbox", i.e. a space where investors, entrepreneurs and researchers can play with the tools available to them.

But some of the people who founded the InterHop association, which "promotes and develops the use of free and open-source software for healthcare", along with other digital activists, have highlighted a serious contradiction within this system. While France wanted to build French sovereignty based on disruptive innovations and unicorns[94], the GIP-PDS specifications are such that the solution chosen to host health data centrally is Microsoft Azure. This choice was made because, at the time, Microsoft Azure seemed capable of providing the required functionality and security for such sensitive data more quickly than other solutions. The GIP-PDS's choice of Microsoft Azure technology was strongly criticised by InterHop[95] , which, along with other claimants, referred the matter to the French Council of State. The order issued by the interim relief judge, no. 444937, on 13 October 2020, takes note of the intention announced by the Government in the course of this legal action to adopt, as quickly as possible, measures to eliminate any risk, such as the choice of a new subcontractor, or recourse to a licensing agreement, as suggested by the National Commission for Information Technology and Civil Liberties (CNIL). The interim relief judge also ordered the GIP-PDS to ensure that Microsoft implements appropriate technical and organisational measures to guarantee maximum protection of the rights of data subjects. Lastly, he ruled that there was an important public interest in allowing the continued use of health data for the purposes of managing the health emergency and improving knowledge of SARS-CoV-2 and, to this end, in allowing recourse to the technical resources, unrivalled to date, available to the GIP-PDS through the contract signed with Microsoft. A European hosting solution, initially planned for 2022, was eventually postponed until 2025[96].

1.4 A regulatory vision for protective sovereignty

The criticism levelled by certain data managers, taken up and added to by many players including the CNIL and the Council of State, has contributed to the emergence of a second notion of sovereignty, which can be called legalistic and protective. It is defended, for example, by Senator Catherine Morin-Desailly and by the cross-party parliamentary committee set up in June 2020 on the theme of "Building French and European digital

---

92 Bothorel E., *Pour une nouvelle ère de la politique publique de la donnée*, report of the parliamentary mission chaired by Éric Bothorel submitted to the Prime Minister on 23 December 2020, 216 p.

93 Mazzucato, M., (2013), *The Entrepreneurial State: Debunking Public vs. Private Sector Myths*. London: Anthem Press, 288 p.

94 See: https://www.elysee.fr/emmanuel-macron/2017/09/20/discours-d-emmanuel-macron-devant-la-72e-assemblee-generale-des-nations-unies.

95 See: https://interhop.org/2020/12/14/stophealthdatahub-donnees-de-sante-en-otage-chez-microsoft.

96 Statement by the director of the GIP-PDS in September 2022 to media outlet Tic Pharma. See: https://www.ticpharma.com/story/2044/stephanie-combes-devoile-le-programme-de-rentree-du-health-data-hub.html

sovereignty", for which MP Philippe Latombe is rapporteur. According to this perspective, it is not a question of reconsidering the plan to make the most of the treasure trove of health data, nor of reconsidering the creation of a GIP-PDS, but of protecting the data itself (and not the markets) against appropriation from outside the European Union. This data are the personal attributes of people residing in France, and must not be transferred to non-European foreign servers, at the risk that all the protections offered by the GDPR, which applies only in Europe, will be useless. Here, sovereignty encompasses the ability to protect individuals' personal data from unlawful use, firstly under French law and then under European law.

Here again, to promote this sovereignty, the State plays a crucial role, but one that is very different from the previous one: of course, it must create a GIP-PDS, in particular to advance science, but at the same time it must promote the creation of a 'Sovereign Cloud', i.e. one whose infrastructures are physically located in Europe so that European legislation applies. It must establish GIP-PDS governance that is more attentive to the criticisms levelled by various national players in the digital ecosystem, in particular the National Health Insurance Fund for Salaried Employees (CNAMTS), which currently hosts a significant proportion of the data via the SNDS and as such benefits from experience that it proposes to share, while not necessarily seeking to centralise all the data. Finally, the State must establish laws to protect this data and punish illegal use.

1.5 A European vision based on the notion of strategic autonomy

1.5.1 The Gaia-X association

Outside the institutional framework of the European Union, but within European borders, the Gaia-X association was created in 2021 on the initiative of France and Germany. It currently comprises more than 350 companies and organisations, including GAFAM[97].

Its aim is not to create a European digital giant to ensure real European economic sovereignty, but to encourage the creation of a networked software federation in Europe capable of connecting cloud service providers and data owners in an environment of trust, and stimulating the creation of new common data spaces in various fields, including healthcare, in compliance with strict rules on portability, interoperability, data self-determination and security. The Gaia-X association defines the concept of "data sovereignty" as "the ability of data market participants to exercise self-determination with regard to the exchange and sharing of data" and "to make educated choices about services that adhere to specific technical specifications and European or national regulations of their choosing"[98]. Gaia-X would thus enable a company located in one EU Member State to find a cloud hosting solution in another Member State, to use computing power in a third State, using a data management interface that could be located in a fourth State, while complying with the security and data protection standards in force in Europe[99].

To achieve these objectives, the association will award operators different levels of labels, depending in particular on the guarantee they offer that data processing and services will

[97] However, only European members can sit on the board of directors: https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html

[98] See: *La revue européenne des médias et du numérique*, n°54 bis- 55 Autumn 2020 [https://la-rem.eu/2021/01/un-chiffre-ou-deux-n54bis-55-automne-2020?action=genpdf&id=15339].

[99] See: *La revue européenne des médias et du numérique*, n°54 bis- 55 Autumn 2020 [https://la-rem.eu/2021/01/un-chiffre-ou-deux-n54bis-55-automne-2020?action=genpdf&id=15339].

be located in Europe and their immunity to non-European laws. The most demanding label is planned for health data.

Although the association refers in its presentations to the notion of digital sovereignty, the inclusion of non-European players in a project with a European vocation raises questions about the relevance of the choice of this term. The association seems closer to the French position, as stated by Henri Verdier, ambassador for digital issues: "Many people see the question of sovereignty as a question of hegemony. But we see it as a question of strategic autonomy"[100]. This expression first appeared in the vocabulary of the European Union in 2013 in relation to defence[101], and was taken up again in 2020[102] in a broader sense and applied in particular to digital technology.

1.5.2 The European health data space

At the European institutional level, with regard to health data, the draft regulation for a European health data space published by the European Commission in May 2022[103] is an example of a European digital tool based on the notion of strategic autonomy. In its opinion of 12 July 2022 on this draft, the European Union's Data Protection Board stressed that the data should be hosted in Europe[104].

1.6 Ethical tensions between HDP visions of sovereignty and autonomy

There are therefore at least three different concepts of sovereignty. Supporters of a liberal vision of sovereignty criticise those who defend a regulatory vision of sovereignty for slowing down the progress of French companies and, in so doing, causing them to lose their place in international competition; they fear that regulations will stifle start-ups without controlling the American multinationals which, due to their might, are able to circumvent the rules. On the other hand, regulators criticise liberals for misinterpreting the "trickle-down effect" they expect from the creation of unicorns, on the grounds that, while unicorns sometimes accumulate huge market capitalisations, they produce minimal added value, and always run the risk of these same capitalisations being quickly sold off to competing foreign companies, particularly non-European ones.

The concept of strategic autonomy as implemented by Gaia-X favours decentralisation and federation. It avoids creating new monolithic players and allows work to proceed on the basis of what already exists. But is it sufficiently demanding and does it not take risks by admitting companies from outside the European Union to the association? Inspired by compliance law[105], strategic autonomy according to Gaia-X, which implies the internalisation of principles by the company, excludes an external regulator: the Gaia-X architecture is self-monitoring. However, wouldn't a non-European operator that has set up a company incorporated in a Member State be able to benefit from the most demanding label if it succeeds in integrating into a structure governed by European law that complies with the criteria, even though its independence with regard to the legislation of the foreign country to which it belongs may still be questioned?

---

[100] Renaissance numérique, (2022), "*Rapport sur la souveraineté technologique européenne*", p. 16. [https://www.renaissancenumerique.org/publications/la-souverainete-technologique-europeenne/]

[101] European Council of 19 and 20 December 2013

[102] European Council of October 2020

[103] European Commission, "Proposal for a regulation - The European Health Data Space", 3 May 2022 [https://health.ec.europa.eu/publications/proposal-regulation-european-health-data-space_fr].

[104] See: https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-032022-proposal_en

[105] Frison-Roche M-A., (2019), "*L'apport du droit de la compliance à la gouvernance d'Internet*", Report commissioned by the Minister for the Digital Economy, 134 p.

Hence the importance, prior to the creation and certification of an HDP, of adopting a multi-disciplinary approach, with experts in healthcare, IT, law and the human and social sciences, to consider the sovereignty issues of this platform and anticipate the risks of its sovereignty being compromised (see Recommendation 9).

However, it is not the future of companies that is the main issue. The ethical value that can be attributed to these visions of sovereignty applied to HDPs can only relate to their relationship to the common good and to the specific principles of bioethics: beneficence and nonmaleficence, promotion of and respect for autonomy, justice and equity, to which can be added the principles of explicability and transparency, specific to the ethics of AI (see section 2 of the introduction). In this respect, the State is expected to ensure, via HDPs, that public health is improved, that it is fair and equitable, and that the health data of every citizen is protected.

An initial ethical tension between the competing visions of sovereignty concerns progress in medical research and care, and therefore the principle of **beneficence**, which presupposes facilitating access to health data and therefore its exchange, and would tend to favour efficiency and rapid performance according to the liberal and entrepreneurial vision of sovereignty which led in particular to the choice of Microsoft Azure for the GIP-PDS.

A second ethical tension concerns the risk of non-European companies taking control of our health databases, a national asset, either by taking control of the HDPs or by using their data. This could seriously undermine the notions of **justice** and **equity** that currently prevail in the French healthcare and social protection system if, despite current regulations in France or Europe, one or other of these companies were to profile individuals on the basis of their health data and then offer them different insurance policies on the basis of this profiling (see Recommendation 10).

A third ethical tension concerns the principles of **explicability** and **transparency**, which could be jeopardised if functions are outsourced due to the inability of the French State or Europe to exercise extraterritorial rights over a foreign partner, as is currently the case, with the exception of the General Data Protection Regulation (GDPR), contrary to what the United States may claim. This risk, like the threat of takeover by a private company, would tend to favour the legalistic and protective vision of sovereignty. But regardless of the model of sovereignty chosen, it is essential to include **reversibility** clauses on the role played by a particular company to ensure that i) the data is stored in a non-proprietary format and can be easily transferred to users, particularly researchers, at minimal cost, ii) the algorithms are developed using standard technologies that are independent of the hosting solution, iii) the elements of the platform are partly reusable (architecture, deployment scripts, etc.)[106] (see Recommendation 11).

A fourth ethical tension concerns the principle of **nonmaleficence** in the face of risks associated with exploiting health data, such as the development of harmful medications targeted at vulnerable groups of people, and the risk of cyber-attacks on HDPs. Faced with this, it is the architecture and infrastructure of HDPs that are at stake, without this leading a priori to favour a particular vision of digital sovereignty. Particular vigilance is required during maintenance operations on IT systems serving HDPs and involving non-European players. These particular contexts may require data to be stored temporarily, increasing

[106] See: https://www.health-data-hub.fr/sites/default/files/2021-05/Etude%20de%20r%C3%A9versibilit%C3%A9%20de%20la%20plateforme%20technologique%20%E2%80%93%20Novembre%202019.pdf;

their vulnerability to cyber-attacks and unauthorised access, or even the export of health data (see Recommendation 12).

2. Forms of valuation of health data platforms

As stated in section I.1, the data is not transferable, but it is possible to sell a right of access to it. However, all HDPs are encountering certain difficulties in establishing a pricing structure for their use. These obstacles can be explained by the fact that the service for which a fee is charged, i.e. access to large health databases, is in its infancy and is caught in a conflict between two different "ways of valuing" data. This concept, highlighted by Boltanski and Esquerré[107], refers to conventional forms of valuation, specific to the societies in which they are found, but sufficiently stabilised to constitute "a collective resource to which agents can refer when they have to navigate the world of objects [to attribute a value to them]".

However, data is not an object in the strict sense of the term. It is an intangible asset, as information that is non-transferable on the one hand, and 'non-rivalrous' on the other, meaning that it can be consumed simultaneously by several people without any direct loss of the asset (several people can work simultaneously on the same database without damaging it), which is impossible with material objects. However, the intuition behind the notion of "methods of valuation" remains useful for our reflection. The health databases to which platforms sell access can be considered in two different ways.

2.1 Valuation based on the cost of creation and maintenance

Firstly, from a so-called "standard" perspective (to use Boltanski and Esquerré's term), databases are viewed from the perspective of the work and investment required to create and maintain them. They are based on the collection, storage and implementation of technological building blocks that may be supplied by HDP partners. The issue is therefore to estimate the cost of these services, for example by means of internal cost accounting within the organisations, and possibly to add a share of added value in order to estimate their value and their access price. In practical terms, it has been observed that remuneration for the various functions and services of an HDP is based on various pricing mechanisms: entry price, copyright on the final product, lump-sum price, free cooperation to create patents.

2.2 Valuation based on expected future profits

But, secondly, a database can be viewed from alternative approach to valuation, which Boltanski and Esquerré call the "asset" form, in which the databases created are viewed as shares of capital that make it possible to produce a future value much higher than the present one by means of financial mechanisms for the resale of a stake in a company. In this case, the expected future valuation is factored into the price calculation. To borrow from the characteristics identified by a benchmark of health database valuation models, commissioned by the Aviesan alliance[108], they have a "capacity to generate value", they involve a risk - and therefore an opportunity - associated with the partnership between the platform holding the data and the party using it, or they are the subject of agreements that provide a greater or lesser incentive for those involved in the project to invest. The value

[107] Boltanski, Luc, and Arnaud Esquerre. *Enrichissement: une critique de la marchandise*. NRF essais. [Paris]: Gallimard, 2017.

[108] The perspective of the "asset" form of health databases is the one adopted by the Aviesan alliance in its benchmark of different "health data valuation models". See: https://cvt.aviesan.fr/outils/enjeux-lies-aux-donnees-de-sante/.

attributed to the data is determined by mechanisms such as the creation of a joint venture between the platform and the company developing a project, or the sharing of a percentage of the latter's revenues.

2.3 European and French debate on the two forms of valuation

These two methods of valuation have both been supported by powerful institutions which legitimise them in equal measure. To summarise, the first model is supported by the French healthcare system, centred on university hospitals and public research institutions, while the second is supported by players such as the Banque Publique d'Investissement and its *Deeptech*[109] plan, as well as by numerous start-ups.

2.3.1 Valuation under the European data governance regulation

The European regulation of 30 May 2022 on data governance (Data Governance Act[110]) opts for the first of the two options to facilitate the right of access to data, by setting low fees. The Act includes provisions on the general framework for the provision of different types of data, including personal data, by public sector bodies including the State, regional or local authorities or bodies governed by public law. Various guarantees are provided, including anonymisation. As the aim is to promote access to this data, the level of fees paid for the re-use of this data will be calculated solely on the basis of the "*costs associated with implementing the procedure for re-using the categories of data" made available (Article 6.* 5). As such, the real value of accessing and using such databases will not be taken into account, since the costs retained will essentially be those of supplying and disseminating the data, anonymisation or other forms of data preparation, and maintenance of the secure processing environment.

2.3.2 How should health data platforms in France be funded?

Based on the observation that current funding for platforms, which is most often based on non-recurring future investment programmes, is insufficient and too uncertain, the Prime Minister has entrusted the Strategic Committee for Health Data with the[111] task of analysing and examining the relevance of our current tools for regulating and funding health products, with a view to formulating recommendations by the summer of 2023.[112] The aim is to fund HDPs on a permanent basis at a level sufficient to ensure that they do not require additional revenue. At the same time, a fee structure should be established for certain players, in the case of public-interest missions, that is affordable so that they have relatively easy access to it. These two approaches are tantamount to viewing health data from the standard valuation perspective (see Recommendation 13).

2.3.3 The draft European data regulation

On 23 February 2022, the European Commission presented a legislative proposal, the Data Act (DA), the aim of which[113] is to ensure better distribution of the value arising from

---

[109] See: https://www.bpifrance.fr/nos-actualites/plan-deeptech-3-chiffres-2-ans-un-seul.

[110] See: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868. Published in the Official Journal of the European Union on 23 June 2022, the DGA will come into force in September 2023.

[111] Created by an order dated 29 June 2021. See: https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043850566

[112] See: https://www.gouvernement.fr/communique/mecanismes-de-regulation-et-de-financement-des-produits-de-sante

[113] CNIL, "*Stratégie européenne pour la donnée : la CNIL et ses homologues se prononcent sur le Data Governance Act et le Data Act*" [https://www.cnil.fr/fr/strategie-europeenne-pour-la-donnee-la-cnil-et-ses-homologues-se-prononcent-sur-le-data-governance].

the use of personal and non-personal data between the players in the data economy, particularly in connection with the use of connected objects and the development of the Internet of Things (IoT). This point concerns health data collected by all connected medical devices, which can be fed into health data platforms. This objective of the DA is very much in line with our recommendation on value sharing (see Recommendation 14).

2.4 Vigilance regarding potential conflicts of interest

Insofar as health data platform valuation methods may involve private interests alongside the public interest, they are not immune to the risk of conflicts of interest. [114] This calls for vigilance in appointments to positions of responsibility and non-competition clauses for public HDP managers (see Recommendation 15).

3. Recommendations

- **Recommendation 9**: Require a multidisciplinary approach to the creation and certification of an HDP, with experts in healthcare, IT, law and the human and social sciences, in order to anticipate the risks of sovereignty being compromised.
- **Recommendation 10**: In international partnership contracts involving health data, ensure that there are clauses guaranteeing that non-European players comply with the fundamental principles of the GDPR, the Data Governance Regulation, and the future European Data Act, to protect sensitive data.
- **Recommendation 11:** Systematically provide for specific transparency, explicability and reversibility clauses, for companies, particularly non-European ones, allowing in particular data transfers at minimal costs.
- **Recommendation 12**: Monitor the conditions of access to health data and temporary export during the maintenance of IT systems serving HDPs involving non-European players.

Valuation of data

- Recommendation 13:: Promote the funding of HDPs on the basis of their investment and operating costs, and pricing adapted to different users, in particular for scientific research in the public interest.
- **Recommendation 14**: Encourage companies that achieve financial success partly thanks to datasets provided by public health data platforms to share part of their profits with the latter, by voluntarily signing a charter committing their reputation.
- **Recommendation 15**: Ensure a high degree of independence between the management teams of public health data platforms and those of user companies in order to prevent conflicts of interest.

Box 4: The issue of sovereignty over radiology HDPs

Following in the footsteps of private clinics, nursing homes and, more recently, biology platforms, which were largely acquired by financial groups from international investment funds, radiologists in the private sector, with their medical imaging technical platforms, are the subject of massive acquisition proposals by investors, whose attractive financial offer conceals numerous uncertainties and risks.

[114] See: https://www.lemonde.fr/planete/article/2019/12/24/donnees-de-sante-conflit-d-interets-au-c-ur-de-la-nouvelle-plate-forme_6023918_3244.html

The first of these risks concerns the **lack of transparency in the structuring** of many of the companies that acquire professional platforms, with three levels of organisation: SEL (Sociétés d'Exercice Libéral), SELAS (Sociétés d'Exercice Libéral par Actions Simplifiées) and financial holding companies that are partners in SELAS, with the result that foreign investors never appear directly in the share capital of the companies with which healthcare professionals will be contractually linked.

This arrangement paves the way for a **twofold lack of transparency in the contracts offered**. The first lack of transparency results from the fact that the Departmental Councils of the French College of Physicians only give their opinion on the articles of association of SELs, and not at all on the related and complex contracts which are signed elsewhere and which are not communicated to them even though they should be. A second lack of transparency concerns professionals, who have no say in the governance, management or control of the financial rights of the radiologist practising in these SELAS.

These contracts, which are multi-layered in terms of content and form, **firstly** result in **the deregulation of regulated professions**, with the proven risks of losing decision-making autonomy and directing activities towards profitable, simple and modelled examinations, as well as a possible infringement of patients' freedom of choice through the signing of exclusivity clauses between the group and certain private clinics or teleradiology platforms (including abroad). This leads to an obvious risk of practices compromising the independence of professionals, which is guaranteed by Article R. 4127-5 of the French Public Health Code, as well as a major infringement of respect for patients' freedom of choice: patients could thus no longer have access to a specialist that they have chosen or that their doctor advises them.

Secondly, there is **a need to clarify the ownership of mass patient imaging data, which may be interpreted, stored and used abroad** (in teleimaging networks) and thus escape any control, despite the fact that these examinations are publicly funded by the national health insurance fund (Assurance Maladie), and given the sensitivity of the personal diagnostic and therapeutic data they contain.

III. CONSENT TO DATA SHARING AND CITIZEN PARTICIPATION IN THE DEVELOPMENT AND GOVERNANCE OF HEALTH DATA PLATFORMS

1. Various forms of consent

The processing of personal health data is generally prohibited, although Article 9 of the GDPR provides for exceptions, in particular where "the data subject has given explicit **consent** to the processing of those personal data for one or more specified purposes" [115].

1.1 Free, informed and specific consent

The definition of consent is given in Article 4(11) and the procedures for obtaining consent are specified in Article 7 of the GDPR[116]. Consent must be **free** (the individual must be able to give his or her consent without constraint), **specific** (one or more purposes must be indicated, and consent must be given separately for each purpose), **informed** (the individual understands the processing that will be carried out on his or her data and guarantees must be given that the individual is aware of the implications of his or her consent) and **unambiguous** (the individual must have explicitly given his or her consent and the data controller must be able to prove this). On this last point, this consent is clearly an **opt-in** mechanism (as long as individuals do not explicitly say yes, it must be considered as no), since silence, pre-ticked boxes or inactivity cannot be considered as consent[117]. Finally, people must be able to **withdraw** their consent whenever they want, and the process for doing so must be easy.

While the characteristics of consent are easy to understand, the implementation of the procedures for obtaining it is less straightforward. The CNIL provides some guidelines for data controllers[118], and the CCNE recently discussed the development of ethical issues relating to consent in healthcare in its opinion 136 (see box 5).

Box 5: CCNE opinion 136

CCNE opinion 136 on "The evolution of ethical issues relating to consent in health care"[119] looked at the issues of the effectiveness of obtaining consent and the increasing complexity of the related ethical issues as a result of the development of new medical techniques. In order to understand the issues at stake, the CCNE took the view that consent, an act of care in its own right, is an evolving and dynamic process which "is not given once and for all, but is developed and may evolve within the framework of a relationship based on mutual trust"[120], including the possibility of withdrawal. In particular, this opinion recommends increasing the role of the trusted representative, to ensure a more respectful approach to the wishes of vulnerable people (who have difficulty expressing their wishes or are unable to decide for themselves). In addition to

[115] See: https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre2#Article9.
[116] See: https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre2#Article7; See also: Recitals 32 (Conditions of consent), 33 (Consent for certain areas of scientific research), 42 (Burden of proof and conditions of consent), 43 (Consent freely given) of the GDPR.
[117] Recital 32: "Silence, pre-ticked boxes or inactivity should not therefore constitute consent".
[118] *Conformité RGPD : comment recueillir le consentement des personnes ?* CNIL, 2018, https://www.cnil.fr/fr/les-bases-legales/consentement.
[119] CCNE, Opinion 136 of 15 April 2021, *L'évolution des enjeux éthiques relatifs au consentement dans le soin*, 51 p.
[120] CCNE, opinion 136, *op. cit.* p.4.

strengthening the role of the trusted representative, the CCNE believes that digital technology and its information tools are the best way to help people express their wishes and remember the consent process, particularly in terms of tracking information and individual pathways.

The requirement for this form of consent in care matters meets a number of ethical imperatives that are well known in bioethics: respect for the right to self-determination (individual autonomy), beneficence, justice, and above all respect for the individual and his or her dignity[121]. In the case of health data platforms (HDP), it is a question of respecting patients' autonomy, in their ability to determine how their data is used.

The difficulty with specific informed consent is that consent is given for a specific purpose and for a specific period of time, which runs counter to the philosophy of HDPs, the function of which is to allow data to be used for subsequent purposes that are not necessarily foreseen. So how can consent be given? Alongside this traditional model of consent, other forms have emerged for foreign biobanks and health data research[122].

1.2 Other forms of consent

In contrast to specific consent, **broad consent** asks individuals to consent to multiple future studies, without knowing the precise details of these studies at the time of the request. Information on the aims, risks and possible benefits is given in a general way. Individuals have less control over the data, since there are no regular exchanges, which means that they have less opportunity to withdraw their consent if they have forgotten that they gave it or if the conditions for using the data change[123].

This is the consent model used by the NIH (National Health Institutes) health data platform for its *All of Us* Research Program[124], which aims to facilitate the development of precision medicine by setting up a large, well-characterised research cohort, while at the same time imposing very precise rules for its use, for example by asking researchers and institutions to provide a plan for sharing genomic data[125]. This is also the case for the British NHS (National Health Service) *100,000 Genomes Project*, launched in 2018, the aim of which is to build up a genome sequence database of around 85,000 NHS patients suffering from a rare disease or cancer, as well as their families, in order to link these diseases with the genes likely to play a role in their onset and development.

The model of **dynamic consent**[126], on the other hand, is not really a form of consent but is based on a personalised online communication platform that facilitates the consent process between researchers and participants. It enables two-way communication and can reinforce the right of research participants to make autonomous choices about their participation in research, improve their understanding of the consent process and foster their engagement in the research enterprise. It also allows consent to be stored. This model is the basis of dynamic specific consent and meta-consent, and presupposes that patients

[121] CCNE, opinion 136, *op. cit.*

[122] Wiertz S., Boldt J., (2022), "Evaluating models of consent in changing health research environments", *Med Health Care Philos*, Jun;25(2):269-280.

[123] Mikkelsen RB, Gjerris M, Waldemar G, Sandøe P. : "Broad consent for biobanks is best - provided it is also deep". *BMC Med Ethics*. 2019 Oct 15, 20(1):71.

[124] See: https://allofus.nih.gov/about/protocol/all-us-consent-process

[125] See: https://sharing.nih.gov/genomic-data-sharing-policy/about-genomic-data-sharing/gds-policy-overview

[126] Budin-Ljøsne I. et al., (2017), "Dynamic Consent: a potential solution to some of the challenges of modern biomedical research", *BMC Med. Ethics*, 18, 4.

have access to digital platforms and a degree of 'digital literacy'. It is in line with the recommendations made by the CCNE in its opinion 136, in which consent is described as an evolving and dynamic process, leading the ethics committee to recommend the use of digital tools in the collection and storage of consent.

In the **dynamic specific consent**[127] model, information is provided to individuals in the format of their choosing, in line with their level of education and interest. Project participants are kept regularly informed of any changes. In this model, MCQs could be used to ensure that participants have fully understood the information they have been given. The genomic platform project, Promise for Engaging Everyone Responsibly (PEER)[128], developed by the Genetic Alliance, requires dynamic specific consent and research participants can use a matrix to indicate which types of access they approve of and which they do not.

Unlike dynamic specific consent, which uses the same consent form for all participants, in the meta consent[129] model people can give specific or broad consent, depending on their personal preferences. Categories are suggested to allow different consent options depending on the fields and formats of the projects. User preferences are managed by the platform and can be changed at any time. Researchers can contact participants by query on the platform. This meta-consent[130] model is used by the GRIIS[131] health data platform developed in Canada: "Instead of consenting to one project at a time, patients and individuals in general would consent to multiple projects with similar characteristics accessing their health data". The main challenge of this model is determining the categories of projects, which could nevertheless benefit from patient participation.

1.3 Post-mortem health data

While the GDPR states that it does not apply to the data of deceased persons, the law of 6 October 2016 for a Digital Republic established a legal framework governing the processing of personal data of such persons. While affirming the principle that rights relating to personal data lapse on the death of the individual, it allows the person concerned to plan ahead for the management of their personal data. For example, under the current Article 85 of the French Data Protection Act, resulting from the Act of 6 October 2016, "any person may define directives relating to the retention, deletion and communication of their personal data after their death. These directives may be general or specific". The data subject may designate a third party to carry out his or her instructions. The risk is that the data subject may be unaware of the existence of these legal provisions. In this respect, we can only welcome the existence of online documents from the Groupement hospitalier universitaire de Paris[132] and the Assistance Publique des Hôpitaux de Paris[133] informing patients of their right to define directives on the storage, deletion and communication of their health data after their death. However, the scope of this right to erasure is open to question, since even though, as we have said, the GDPR does not apply

[127] *Ibidem*.
[128] See: https://geneticalliance.org/registries/promise-for-engaging-everyone-responsibly.
[129] Ploug T., Holm S., (2016), "Meta Consent – A Flexible Solution to the Problem of Secondary Use of Health Data", *Bioethics*, 30 (9), pp. 721 – 732.
[130] See: https://griis.ca/recherche/claret/.
[131] Cumyn A. et al., (2021), "Meta-consent for the secondary use of health data within a learning health system: a qualitative study of the public's perspective", *BMC medical ethics*, vol. 22,1 81.
[132] See: https://www.ghu-paris.fr/fr/lentrepot-de-donnees-de-sante-eds.
[133] See: https://www.aphp.fr/patient-public/vos-droits/protection-des-donnees-personnelles-information-patient

to deceased persons, the national law merely refers to the GDPR. However, the right to erasure provided for in Article 17 of the European Regulation is not an absolute right. It cannot be exercised when the processing is in response to a legal obligation or is necessary for the performance of a public service mission.

Furthermore, under the terms of Article R. 1112-7 of the Public Health Code, health establishments must retain medical records for a period of ten years after the deceased patient's last visit to the establishment. The French Public Health Code[134] organises the communication of medical records to the deceased person's legal successors in a balanced manner that respects both the wishes of the deceased and the preservation of medical confidentiality. Communication of medical records to legal successors is not authorised if the deceased has objected, and such communication is only permitted to a limited extent: to investigate the cause of death, to exercise a right or to defend the memory of the deceased. The beneficiaries will only be authorised to access the elements necessary for the objective pursued.

It is therefore desirable that all health data warehouses or platforms explicitly and clearly inform registered persons of their right to erase this data after their death, specifying the scope and limits of this right (Recommendation 16).

1.4 Ethical issues of consent

1.4.1 Advantages of consent models

- Consent makes the aims of HDP projects transparent, which strengthens public confidence by giving research projects a higher profile.
- In the context of HDPs, informed consent (especially dynamic specific consent) enables participants in research projects to be informed of the aims pursued by researchers and to be able to compare them with their own values and interests, and thus to decide whether or not to become involved.
- Dynamic consent, which uses a platform to store consent, enables patients to keep track of their choices. This is in line with the recommendation made by the CCNE in its opinion 136 (collection and recording of consent).

The advantages of this type of platform are illustrated by the following example. With regard to the *100,000 Genomes Project*[135], which uses broad consent, a study in 2020[136] showed that some of the participants in the study did not grasp the complexities of the project and the types of results it could lead to; for example, 20% of the participants in the "cancer" section questioned could not remember the decisions they had made concerning secondary discoveries.

1.4.2 Limitations of consent models

- Different countries apply different consent models, which poses a challenge for large-scale international research projects.
- Online consent poses a problem of authentication (problem of electronic signatures)[137].

[134] See Articles L. 1117-7 and L. 1110-4 of the French Public Health Code.

[135] See: https://www.genomicsengland.co.uk/initiatives/100000-genomes-project

[136] Ballard L.M., Horton, R.H., Dheensa, S. *et al., (2020),* "Exploring broad consent in the context of the 100,000 Genomes Project: a mixed methods study", *Eur J Hum Genet* 28, 732–741. https://doi.org/10.1038/s41431-019-0570-7.

[137] Kogetsu, Atsushi, and Kazuto, (2022), "Framework and Practical Guidance for the Ethical Use of Electronic

- The "informational" risk - i.e. the completeness of the information - must not be overlooked when seeking to obtain specific consent. As Mikkelsen et al.[138] point out, rather than the objectives of HDP research projects, it is the way in which HDP data is secured that may encourage people to consent and deposit their data. Similarly, a feeling of trust in the project leaders is fundamental to obtaining consent.
- The cumbersome nature of setting up platforms for dynamic consent must be taken into account: the cost may be to the detriment of the research, and the time needed to obtain consent considered too great by researchers before a project starts.
- As in the case of cohort surveys, there is a feeling of fatigue with dynamic consent due to the routinisation of the click (consent fatigue): by seeking people's consent too frequently, you tire them out[139].

1.4.3 Tensions and ethical issues

The collection of health data for HDPs once again highlights the tension between data protection, respect for privacy (free and informed consent of patients) and contribution to the common good (medical progress and improvement of public health).

This consent must be weighed against the possibility of patients waiving the confidentiality of their data, as emphasised in CCNE opinion 136[140].

The request for consent highlights an issue of equity. Researchers[141] have observed that the willingness to share one's data and to give broad consent to research projects is unevenly distributed according to social group (ethnic origin, gender, socio-economic level), especially with broad consent, which is similar to the altruism with regard to data discussed below (section III.4). Several consequences are worth mentioning:

- There is a risk of representation bias in HDPs, leading to an increase in health inequalities. This was addressed in section II.2.
- Individuals who may not fully understand the terms and purposes of research projects could give broad consent, even to dubious studies, thereby putting their personal data at risk.
- In the case of dynamic consent, people who do not have the equipment to access online platforms or who have low digital literacy could be prevented from giving their consent, potentially depriving the community of which they are a part of the potential benefits of the research.

Consequently, dynamic consent must be encouraged by ensuring confidence in HDPs through regular and transparent information, without being overly burdensome (see Recommendation 17). Care must be taken not to inundate patients with information, and to encourage digital literacy and support from trusted representatives who could be digital health assistants[142] (see Recommendation 18).

Methods for Communication With Participants in Medical Research", *Journal of medical Internet research* vol. 24,4 e33167.

[138] Mikkelsen R.B., Gjerris M., Waldemar G., Sandøe P., (2019), "Broad consent for biobanks is best - provided it is also deep", *BMC Med Ethics,* Oct 15;20(1):71.

[139] *Ibidem*.

[140] CCNE, opinion 136, *op. cit.*

[141] Wiertz S., Boldt J., (2022), "Evaluating models of consent in changing health research environments", *Med Health Care Philos*, Jun;25(2): Considerations of justice, page 272.

[142] See: Joint opinion no. 141 of the CCNE and no. 4 of the CNPEN, (Jan. 2023), *Diagnostic Médical et Intelligence Artificielle : Enjeux Ethiques*, 58 p.

In addition, for this information to be passed on correctly to patients, a minimum level of training is required for caregivers, or for the trusted representative as digital health assistant. The trusted representative could also receive this minimum training via patient associations. It is therefore crucial to provide training in the technical and ethical issues surrounding health data platforms for caregivers, as well as for patient representatives and trusted representatives who would play the role of digital health assistants (see Recommendation 19).

2. Opt-out

While a number of HDPs expressly request consent to collect and process personal data from individuals (opt-in strategy), others have an opt-out strategy, which does not require explicit consent from individuals: as long as they do not say no, they can be considered to have agreed.

2.1 In France

The opt-out strategy is applied to health insurance data. As stated on the Ameli website[143]: "Apart from a few exceptions, you cannot object to the Assurance Maladie using your data as part of its missions set out by law or for public health reasons. Specific information is therefore provided to you via our specific information notices when you have the right to object". For its part, the French Data Protection Authority (CNIL) explains in a note[144] what applies to the SNDS warehouse (and therefore to the GIP-PDS): "All individuals have the right to object if they do not wish their data contained in the SNDS to be used for research purposes. They may not, however, object to the processing of data necessary for the performance of the tasks of State services and certain public establishments such as, for example, monitoring an epidemic or health surveillance".

The opt-out strategy is also the strategy adopted by *Mon espace santé*[145], unlike the previous initiative of the *Dossier Médical Partagé* (DMP), which it replaces from 1 July 2021, and which proposed an opt-in strategy. Created in the spring of 2022, "*Mon espace santé* enables anyone affiliated to the French national health insurance fund (Assurance Maladie) to store and access their health data in complete confidence and security. (...) This digital health space is activated online, using a temporary secret code received by email or post. However, activation is by no means compulsory. If you do not act within 6 weeks of receiving the temporary password, your account will be created automatically"[146]. In February 2023, 98% of policyholders had a *Mon espace santé* profile, 7.92 million accounts had been activated and 26% of users had completed their medical profile.[147]

2.2 In the United Kingdom: "National data opt-out"

In the UK, the national data opt-out[148] is a national data service that was introduced on 25 May 2018. It allows patients to opt out of having their confidential information used for purposes other than their individual care and treatment, whether for research (e.g. medicines for rare diseases) or planning (improving health and care services). The national data opt-out covers confidential patient information collected about care in the UK: publicly

[143] See: https://www.ameli.fr/assure/protection-donnees-personnelles
[144] CNIL, "SNDS : Système National des Données de Santé" [https://www.cnil.fr/fr/snds-systeme-national-des-donnees-de-sante].
[145] See: https://www.monespacesante.fr/.
[146] See: https://www.aide-sociale.fr/mon-espace-sante-suppression-compte/#.
[147] See: https://esante.gouv.fr/strategie-nationale/mon-espace-sante accessed on 13/02/2023
[148] See: https://digital.nhs.uk/services/national-data-opt-out.

funded, commissioned or coordinated health and adult social care, and private care provided under the National Health Service (NHS).

All NHS organisations must provide information about the type of data they collect and how it is used. Data release registers are published by NHS Digital and Public Health England, showing what data they have shared with other organisations.

If patients agree to their health data being used for purposes other than their individual care and treatment, for research and planning, they do not have to do anything. They can, however, view or change their opt-out choice from national data at any time using an online service or via a click in the NHS application. The NHS[149] online site lists the various cases where, even if the patient refuses, their health data may nevertheless be used.

3. Opt-in/opt-out ethical tensions

According to the European Commission's report on health data in the light of the GDPR[150], some studies[151,152] show that the opt-in consent model is considered to be a more reliable data-sharing practice, while other studies[153] show that the opt-out approach is also acceptable if certain conditions are met.

The ethical tension between the two approaches, between respect for privacy and serving the general interest, is again apparent. The opt-in approach, which is based on free and informed consent, favours respect for personal data, possibly at the expense of feeding the HDPs, while the opt-out approach favours enriching the HDPs, thereby reducing data bias and making research more effective, but possibly at the expense of informing patients about the use of their personal data.

The opt-in approach is more transparent. However, free and informed consent presupposes that patients have sufficient medical information and a minimum level of digital skills, in the case of dynamic consent on a platform, to give their consent, or that they have the option of using a trusted third party to help them do so. Without this, there is an issue related to equity.

However, this transparency may be perceived as burdensome, or even anxiety-provoking, by patients, who would be regularly informed of the launch of new research projects and asked to give or withhold their consent, depending on the frequency of this communication.

It is therefore important to take into account the tension between respecting privacy by providing the information needed to clarify consent and respecting the peace of mind of patients who do not want to be solicited too often.

---

[149] See: https://www.nhs.uk/your-nhs-data-matters/where-your-choice-does-not-apply/.

[150] European commission, *Assessment of the EU Member States' rules on health data in the light of GDPR*, (2021), https://ec.europa.eu/health/system/files/2021-02/ms_rules_health-data_en_0.pdf

[151] Karampela M., Ouhbi S., & Isomursu M., (2019), "Connected Health User Willingness to Share Personal Health Data: Questionnaire Study", *Journal of Medical Internet Research*, 21(11).

[152] Stockdale J., Cassell J., & Ford E., (2019), "Giving something back: A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland", *Wellcome open research*, 3, 6.

[153] Skovgaard L., Wadmann S., Hoeyer K., (2019), "A review of attitudes towards the reuse of health data among people in the European Union: The primacy of purpose and the common good", *Health Policy*, 123 (2019) 564–571.

4. Altruism in relation to health data

4.1 A new form of consent for the common good

Altruism with regard to data is presented as an alternative to the default model of data collection (opt-out), and is closer to the concept of broad consent already discussed (section III.1.2). In its February 2021 report[154], the European Commission mentions several citizens' initiatives that have emerged in a number of countries as part of a bottom-up approach. These involve the sharing of health data, organised by cooperatives owned by individuals, such as Salus-Co-op (see Appendix 4.12), enabling them to become active participants in the scientific field. The voluntary participation of individuals in cohorts such as Constances and UK Biobank (see section I.2.4 and appendices 4.9 and 4.10), who make their personal health data available to researchers over and above the content of standardised national HDBs, may also be considered as altruism with regard to data.

Eric Salobir, a priest and Vatican expert on new technologies, defends altruism with regard to data, arguing that "protecting data without recognising its value it is only half the battle. Digital sovereignty also lies in making good use of the resources at our disposal. [...] If we want to put data at the service of the general interest, we need to be more selective about user profiles and more pragmatic in finding ways of financing the management of this data. This is what the data-altruism model proposes"[155].

While altruism with regard to personal health data is in principle a virtue in terms of contributing to the general interest, public health and progress in medical research, and while it is encouraged in the creation of HDPs based on cohorts and by patient associations[156], we must nevertheless be wary of the possible misuse of this sensitive personal data in relation to the initial intentions of the volunteers who make it available. Hence the importance of the specific legal framework currently being developed.

4.2 Altruism in the European data governance regulation

In order to promote the availability of data and the re-use of certain categories of protected public sector data, and to create a reliable environment to facilitate their exploitation for the purposes of research and the creation of new services and innovative products in the general interest, the European Commission has adopted the Data Governance Act (DGA)[157]. This text introduces the notion of data altruism, defined as the "voluntary sharing of data based on consent by data subjects to process personal data pertaining to them, or permissions of other data holders to allow the use of their non-personal data without seeking or receiving a reward"[158].

This voluntary sharing is for purposes of general interest, such as public health, and concerns both individuals and businesses. The regulation provides for the creation of two

---

[154] European commission, *Assessment of the EU Member States' rules on health data in the light of GDPR*, (2021), https://ec.europa.eu/health/system/files/2021-02/ms_rules_health-data_en_0.pdf, p. 113.

[155] Salobir E., quoted in *La Croix*, (Débat), "*Faut-il partager ses données au nom de l'intérêt général ?, Partager ses données sans craindre pour sa vie privée*", 28 June 2022.

[156] One example is the multi-party agreement signed by France Assos Santé with the GIP-PDS, Santé publique France and Sanoïa. See: https://www.santepubliquefrance.fr/presse/2022/les-associations-s-engagent-pour-l-ouverture-des-donnees-de-sante-a-la-recherche-d-interet-public-une-convention-multipartite-entre-france-assos

[157] See: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868. Published in the Official Journal of the European Union on 23 June 2022, the DGA will come into force in September 2023.

[158] See: https://www.europarl.europa.eu/doceo/document/A-9-2021-0248_FR.html.

new players in data collection, one a service provider operating within a commercial framework and the other an altruistic organisation.

The first player, known as a "data intermediation service provider" (Articles 10 and 11), is required to notify the Member State in which it has its principal place of business - unlike the data broker presented in section I.2.5. These data intermediation service providers act in return for remuneration between the data holders and the potential users of these data, including by providing the technical or other means necessary to enable these services to be provided. This new regime for data intermediation service providers defines the conditions associated with the provision of these services and prohibits them from using data for any purpose other than making it available to users, thus reiterating the principle of limiting the purposes for which personal data is processed. The aim is to establish trust and guarantee the neutrality of these organisations. Examples of this type of structure outside the healthcare sector include the Dawex data exchange platform (see Appendix 4.11) in France and Deutsche Telekom's Data Intelligence Hub[159] in Germany.

The creation of this second new data actor is part of States' development of national policies in the area of data altruism (Article 16). If, in this context, individuals agree to voluntarily make available, out of altruism, personal data concerning them held by public sector bodies, the altruistic data organisations that will be created on this basis to collect this data for general interest purposes must be independent and non-profit-making**.** To be recognised as such, these altruistic data organisations will have to be entered in a register to be kept at national and European level, which will enable them to be recognised throughout the European Union. Article 21 lays down very strict guarantees regarding the people from whom the personal data originates. Altruistic organisations must not use data for purposes other than those of general interest for which the data subject or data holder authorises the processing. They must not use deceptive commercial practices to solicit the supply of data. They shall provide tools for obtaining consent from data subjects or authorisation to process data made available by data holders, as well as tools for easily withdrawing such consent or authorisation.

The regulation also provides for the creation of a consent form (Article 25), which will be valid throughout the European Union as a model consent form for data sharing and re-use, in order to increase transparency for data subjects and build the confidence needed to encourage individuals and businesses to send their data to these organisations.

4.3 Vigilance with regard to data altruism

The CNIL and its European counterparts have stressed the need to ensure that the two European texts, the Digital Governance Act and the Digital Act, are consistent with the GDPR and have called for "intelligent governance revolving around the data protection authorities in order to ensure the efficient and effective application of the various legal frameworks and to ensure that they are legible for the individuals and economic players concerned"[160]. However, even though it is clearly specified that the terms of altruistic data consent must comply with the GDPR and that in the event of conflict with the Data Governance Regulation, the GDPR takes precedence, it is not easy to enforce one's rights, or even to know whether they are being respected. Once a person has given their consent, they seem to have no visibility of what happens to their data, which we find regrettable (see Recommendation 17).

[159] See: https://dih.telekom.com

[160] CNIL, "*Stratégie européenne pour la donnée : la CNIL et ses homologues se prononcent sur le Data Governance Act et le Data Act*" [https://www.cnil.fr/fr/strategie-europeenne-pour-la-donnee-la-cnil-et-ses-homologues-se-prononcent-sur-le-data-governance].

Furthermore, altruism with regard to data can lead to poorly representative data, since the content of the platform that collects it depends on the goodwill of the people who feed it. Although this risk of data bias for HDPs (see section I.2.1) is not specific to data altruism, it is certainly accentuated in this case. To correct for these biases in the representativeness of the population, a 'control sample' can be set up, as proposed by the Constances[161] cohort, by asking non-participants in Constances to provide their data from the Assurance Maladie and the Caisse nationale d'assurance vieillesse, which will be processed anonymously, in order to achieve better representativeness.

Finally, there are two risks associated with altruism with regard to data, as highlighted by the HTF-Sopra Steria Next report[162]: (i) greenwashing (or bluewashing): a company could put forward an altruistic approach to data in order to mask harmful behaviour that it may have engaged in and which has been highlighted by the media and (ii) deception: the European Consumer Organisation reiterated its concern "about the way in which a vague definition of altruism [...] could allow companies to abuse vague and altruistic motives to encourage consumers to share their data"[163], with various forms of phishing being conceivable. Hence the need for transparency.

The use of data altruism depends on three factors: (i) donors' trust in those who will have access to their data, (ii) the greater or lesser degree of sensitivity of the data, which is very high in the case of HDPs, and (iii) the extent to which the data can contribute to the public good. Among the obstacles and reservations is the fact that the data may be used for commercial purposes or by government organisations.

Finally, this model of data altruism does not take sufficient account of the real value of accessing and using HDP data, and the cost to the latter, as, for example, for the AP-HP HDP. Furthermore, there is no provision for any benefit in return for people who have altruistically given access to their data, for example in terms of research results. On this point, the HTF-Sopra Steria report[164] states that "a data altruism model may include (i) funding mechanisms to ensure the implementation and maintenance of sharing infrastructures (ii) incentives to participate in the models, whether direct or indirect". It adds that: "While the DGA already provides that trusted third parties can set up systems to grant access to the data they hold in return for fees, thought should be given to the introduction of direct or indirect incentive systems for data contributors"[165]. On this last point, we recommend the recognition of health data contributors (see Recommendation 14).

5. Towards a collaborative ecosystem for health data platforms

We have identified a number of conditions that the design and management of health data platforms must meet in order to ensure a health ecosystem that respects the ethical principles set out in this opinion. They are in line with the summary of the TEHDAS[166] citizens' consultation, which concludes that "the conditional beneficence of citizens with

[161] See: https://www.constances.fr/espace-volontaires/representativite.php

[162] HTF – Sopra-Steria Next report, (2022), "*Data-altruisme, une initiative européenne. Les données au service de l'intérêt général*", Human Technology Foundation and Exploratoire Sopra Steria Next report, 28.

[163] L. Bertuzzi, "Data Governance: new EU law for data-sharing adopted", *Euratic*, 1 December 2021. Quoted in the HTF-Sopra Steria report, *op. cit.* p 41.

[164] HTF-Sopra Steria report, *op. cit*.

[165] *Ibidem*.

[166] TEHDAS: Towards European Health Data Space, (2022), *Le Débat des Données, une consultation citoyenne en ligne sur la réutilisation des données de santé – Interim report, 36 p.*

regard to the re-use of health data requires a well thought-out framework, paying sufficient attention to its ethical, legal and societal dimensions".

5.1 Guaranteeing the common good

The ethical tension between the protection of health data, respect for privacy (free and informed consent of patients) and the contribution to the common good through the sharing of data (medical progress and improvement of public health) highlighted in section III.1.4 raises the question of defining the boundaries of the common good with regard to the use of health data. One difficulty arises from the fact that this common good may vary according to the information, interests and preferences of individuals. This should lead to citizens being involved in the governance of HDPs in an efficient way. This would enable them to express their views on the consent procedures to be put in place and on their research needs (see Recommendation 21).

This involvement is not self-evident.[167] A number of surveys, including the one carried out by the Roche Foundation in 2021, show that individuals have very unequal access to digital data, and are therefore very unaware of the risks and benefits associated with this data[168]. Isolated citizens will therefore tend to overlook the importance of their involvement in HDPs, in terms of both infrastructure design and governance.

5.2 Encouraging citizen participation in the governance of health data platforms via associations

On the other hand, when patients are organised, for example in a patient association or through an institute, they are much more aware of the importance of the role they are likely to play and are therefore more involved. The example of the Renaloo association for kidney disease patients is very enlightening in this respect. Yvanie Caillé, a founding member, has shown her interest in participating in governance by becoming director of the French National Institute for Health Data. Similarly, one of the governance bodies of the cancer data platform of the French National Cancer Institute (INCa) (Appendix 4.7) is its scientific and ethics committee, which includes patient representatives. Citizen participation in the governance of HDPs is much more effective when it takes place through patient organisations.

More generally, the ethical issues relating to the health ecosystem at the heart of HDPs lead to the development of processes for public consultation and information at all stages of the health data processing chain: upstream, when research projects are defined and decisions are taken on how to leverage value from the research, during the research phase, and afterwards, so that the public can be kept informed of the results obtained (see Recommendation 20).

5.3 Building trust through information, transparency, training and digital support

With regard to patient information, we recommend that a booklet or information sheet be produced and made available in waiting rooms in hospitals and doctors' surgeries, as well as in pharmacies, informing people about how practitioners use their data (see Recommendation 18). This is vital, in particular, for information relating to *Mon espace santé*, which has been in operation since the beginning of 2022, given that caregivers do not inform patients whether they are transferring the data they collect about them to this

[167] Fréderic Graber, *L'inutilité publique ne va pas de soi, Histoire d'une culture politique française*, Paris, éditions Amsterdam, 2022, 208 p.

[168] See: https://fondationroche.org/wp-content/uploads/sites/8/2021/10/fondation-roche_rapport-observatoire-acces-numerique-2021.pdf

space, nor do they ask them for permission to access *Mon espace santé* when it is activated by patients.

It should be remembered that caregivers must be able to provide this information orally, which requires a minimum level of training in the technical and ethical issues involved, as recommended above (see section III.1.4 and Recommendation 19).

The dissemination of clear and reliable information contributes to the quality of interpersonal relations between caregivers and patients, and encourages the sharing of data between patients, caregivers and storage platform administrators.

While patients expect a great deal of transparency from caregivers, it must be stressed that the data collected by caregivers is a reflection of their medical practices, which are visible and accessible to patients. The aim is to establish a relationship of mutual trust between caregivers and patients.

From an ethical point of view, the conditions required for such trusting sharing are honesty, benevolence, non-surveillance and patient autonomy.

This trust is based on technical properties for HDPs, such as security, interoperability, portability, anonymisation or pseudonymisation of data for well-defined uses (see section I.4), sovereignty (see section II.1), and on the guarantee of use in compliance with the GDPR, which safeguards the confidentiality of data and its use for the common good. It is consolidated by an ecosystem of consultation, where transparency of HDP uses prevails, showing the advantages of their design and use, without hiding their limitations and the possible risks they entail. The same transparency is expected for research projects, where the aims pursued must be clearly announced, and the major areas of research should be decided in consultation. The fate of primary-use data fed into HDPs and of data resulting from research work must be made clear, and their valuation must be the subject of consultation. Finally, it should be borne in mind that the development of these HDPs raises the issue of equal access to people's health data, due to inequalities in terms of digital literacy.

6. Recommendations

- **Recommendation 16:** Clearly inform patients that they have the right to specify instructions relating to the retention, erasure and disclosure of their personal data contained in an HDP after their death.
- **Recommendation 17**: Promote a form of consent that preserves the link between the person giving consent and the person receiving it, such that the person can provide consent in confidence to types of projects (and not specific projects) by being informed in a transparent and regular manner about the projects and partnerships that will use their data.
- **Recommendation 18:** Develop information for individuals on the use of their health data and existing HDPs that is adapted to their digital culture through various channels and in the places they frequent (hospitals, pharmacies, private clinics) and offer support from trusted third parties who could be digital health assistants.
- **Recommendation 19**: Create training courses to meet the need for new medical and digital skills relating to HDPs for caregivers, IT specialists, HDP users and digital health assistants.

- **Recommendation 20:** Support the implementation of altruism with regard to health data by regularly and transparently informing people who make their data available about the uses made of their data.
- **Recommendation 21**: Encourage public participation in the governance of HDPs and in the preparation of calls for research projects, in particular through patient associations, and inform the latter before, during and after research projects.

APPENDICES

Appendix 1: Recommendations

In the following recommendations, "Health Data Platform" is abbreviated as HDP.

Data quality and sharing

- **Recommendation 1:** Explain the nature and origin of personal health data collected in HDPs, distinguishing between their primary and secondary uses and, for a given research project, use unbiased datasets, or, where this is not possible, take account of these biases in their analysis, for example, through weighting methods.
- **Recommendation 2:** Ensure that the retention period for the public health data collected is properly calibrated in relation to the requirements of the research, without neglecting the necessary protection of personal data.

Environmental impact of HDPs

- **Recommendation 3**: Evaluate the environmental impact of HDPs and aim for energy sobriety through appropriate choices of data storage, architecture and operating modes.

HDP architecture:

- **Recommendation 4**: Ask public authorities to become more involved in the development of standards and norms for formatting and structuring health data in order to promote better portability and interoperability of HDPs.
- **Recommendation 5**: Carry out comparative evaluation studies between centralised and decentralised approaches to HDPs and their combinations, to ensure secure management of health data. Encourage innovations in federated Artificial Intelligence to inform the debate between centralised and decentralised architectures.
- **Recommendation 6:** Choose HDP architecture solutions that respect local ecosystems and take into account multi-centre research projects that require data distributed across various clinical centres or hospitals, highlighting the benefits of pooling data.
- **Recommendation 7**: Encourage public HDP creators to adopt open standard formats and open source algorithms to enhance data quality and subsequently process data flows, and also enable multi-centre studies, in order to release the innovation potential of all reusers of health data.

Anonymisation:

- **Recommendation 8**: Develop research into alternative methods to data anonymisation and pseudonymisation, in particular homomorphic encryption techniques, in order to make better use of health data.

Sovereignty:

- **Recommendation 9**: Require a multidisciplinary approach to the creation and certification of an HDP, with experts in healthcare, IT, law and the human and social sciences, in order to anticipate the risks of sovereignty being compromised.

- **Recommendation 10**: In international partnership contracts involving health data, ensure that there are clauses guaranteeing that non-European players comply with the fundamental principles of the GDPR, the Data Governance Regulation, and the future European Data Act, to protect sensitive data.
- **Recommendation 11:** Systematically provide for specific transparency, explicability and reversibility clauses, for companies, particularly non-European ones, allowing in particular data transfers at minimal costs.
- **Recommendation 12**: Monitor the conditions of access to health data and temporary export during the maintenance of IT systems serving HDPs involving non-European players.

Valuation of data

- Recommendation 13:: Encourage the funding of HDPs on the basis of their investment and operating costs, and pricing adapted to different users, in particular for scientific research in the public interest.
- **Recommendation 14**: Encourage companies that achieve financial success partly thanks to datasets provided by public health data platforms to share part of their profits with the latter, by voluntarily signing a charter committing their reputation.
- **Recommendation 15**: Ensure a high degree of independence between the management teams of public health data platforms and those of user companies in order to prevent conflicts of interest.

Conditions for a collaborative ecosystem for HDPs:

- **Recommendation 16:** Clearly inform patients that they have the right to specify instructions relating to the retention, erasure and disclosure of their personal data contained in an HDP after their death.
- **Recommendation 17**: Promote a form of consent that preserves the link between the person giving consent and the person receiving it, such that the person can provide consent in confidence to types of projects (and not specific projects) by being informed in a transparent and regular manner about the projects and partnerships that will use their data.
- **Recommendation 18:** Develop information for individuals on the use of their health data and existing HDPs that is adapted to their digital culture through various channels and in the places they frequent (hospitals, pharmacies, private clinics) and offer support from trusted third parties who could be digital health assistants.
- **Recommendation 19**: Create training courses to meet the need for new medical and digital skills relating to HDPs for healthcare staff, IT specialists, HDP users and digital health assistants.
- **Recommendation 20:** Support the implementation of altruism with regard to health data by regularly and transparently informing people who make their data available about the uses made of their data.
- **Recommendation 21**: Encourage public participation in the governance of HDPs and in the preparation of calls for research projects, in particular through patient associations, and inform the latter before, during and after research projects.

Research and innovation

Some of these recommendations relate more specifically to research and innovation. These are Recommendation 2 on the duration of data storage for research purposes, Recommendation 5 on federated artificial intelligence research and Recommendation 8 on alternative methods to data anonymisation and pseudonymisation.

Appendix 2: Members of the working group

Joint working group set up by the National Consultative Ethics Committee for Health and Life Sciences (CCNE) and the National Pilot Committee for Digital Ethics (CNPEN).

Gilles Adda (CCNE&CNPEN)

Thomas Bourgeron (CCNE)

Jacques Bringer (External guest - ERE Occitanie)

Sophie Crozier (CCNE)

Pierre Delmas-Goyon (CCNE)

Emmanuel Didier (CCNE) Rapporteur

Christine Froidevaux (CNPEN) Rapporteur

Fabrice Gzil (CCNE)

Jeany Jean-Baptiste (CNPEN)

Claude Kirchner (CCNE&CNPEN)

Caroline Martin (CCNE&CNPEN)

Jérôme Perrin (CNPEN) Rapporteur

Valéry Ravix (External guest - ERE PACA-Corse)

Camille Darche (editor)

Hanna le Derrien (trainee)

Lucie Guimier (editor)

Anaëlle Martin (editor)

Amélie Turci (trainee)

Appendix 3: Legal risks regarding the transfer of data to the United States

On 13 October 2020, the interim relief judge at the French Council of State rejected a request for the suspension of the Health Data Hub (GIP-PDS)[169]. The request referred to the risk of data being transferred to the United States.

The interim relief judge first ruled on the basis of the contractual stipulations and noted that the Health Data Hub (GIP-PDS) and Microsoft's Irish subsidiary had undertaken, by means of a contract dated April 2020, to refuse any transfer of health data, then stored in the Netherlands, outside the European Union, that Microsoft will not process the Platform's data outside the geographical area specified by the Platform without its approval and that, should access to the data be necessary for the purposes of the operations of the online services and incident resolution carried out by Microsoft from a location outside this area, it would be subject to the prior authorisation of the Platform, which has undertaken not to grant such authorisation. He added that a ministerial order dated 9 October 2020 prohibits any transfer of personal data outside the European Union under this contract. The interim relief judge also asked the Platform to specify in a new addendum that it would not authorise any transfer of data, thereby enabling the order of 9 October 2020 to be incorporated into the contract. Consequently, any transfer of data, even for maintenance purposes, is prohibited.

However, these contractual guarantees did not seem sufficient in the face of US law. The interim relief judge noted that it could not be entirely ruled out that the American authorities, as part of surveillance and intelligence programmes, may request access to certain data from Microsoft and its Irish subsidiary, and that while this risk did not justify suspending the Platform in the very short term, it did require special precautions to be taken, under the supervision of the CNIL.

A transfer request could be made by the United States on two legal grounds: Article 702 of FISA and Executive Order EO 12333, even though the data is hosted in the European Union and the terms of the contract between the Health Data Platform and Microsoft would prohibit it.

Firstly, the interim relief judge noted that the technical measures implemented by Microsoft or likely to be implemented in the near future do not rule out all possibility of that company accessing the data processed under the responsibility of the Health Data Platform, despite the precautions, limiting this risk, associated with the encryption of the data and the storage of the encryption keys used. Microsoft could therefore access the Platform's data.

Secondly, given the existence of a risk, and bearing in mind that the interim relief judge can only order very short-term measures, he asked the Platform to continue, under the supervision of the CNIL, to work with Microsoft to strengthen the protection of the rights of data subjects to their personal data while awaiting a solution that will eliminate any risk of access to personal data by the US authorities, as announced by the Secretary of State for Digital Affairs on the day of the hearing at the Council of State (potential choice of a new subcontractor, recourse to a licensing agreement suggested by the CNIL, etc.). The interim relief judge deemed it necessary to continue using the data for the purposes of managing the health emergency and improving knowledge of SARS-CoV-2.

In addition to this emergency decision, which does not set a legal precedent, reference should be made to the CJEU's Schrems II judgment (C-311/18) of 16 July 2020. The Court, which did not consider the specific question raised by the data collected by the GIP-PDS, as the issue was not at stake, examined the question in general terms.

[169] See: https://www.legifrance.gouv.fr/juri/id/CETATEXT000042444915.

With regard to Article 702 of FISA, it considers that it does not in any way indicate the existence of limitations on the authorisation it contains for the implementation of surveillance programmes for foreign intelligence purposes, nor the existence of safeguards for non-US persons potentially targeted by these programmes.

With regard to the surveillance programmes based on EO 12333, the Court ruled that, based on the evidence before it, the decree did not confer any rights enforceable against the US authorities in the courts.

As regards the two programmes, the Court ruled that the surveillance programmes based on those provisions could not be regarded as being limited to what was strictly necessary.

With regard to the obligation of judicial protection, it also ruled that neither of the two texts ensured an effective remedy before an independent and impartial court in compliance with the conditions set out in Article 47 of the Charter of Fundamental Rights.

Microsoft argued that the GIP-PDS did not fall within the scope of this ruling because, in its view, neither Article 702 of FISA nor EO 1233 could be used as a legal basis for monitoring the health platform for various reasons.[170] This is a significant legal question for which there is no existing legal precedent. Even if Microsoft were right, public opinion and the defenders of individual liberties would still be suspicious and concerned about the American practices revealed by E. Snowden. It is preferable to ensure real autonomy in the area of health data, before waiting for a legal solution, especially as US laws and regulations can always be tightened.

[170] National Assembly, Information Mission of the Conference of Presidents, "*Bâtir et promouvoir une souveraineté numérique nationale et européenne*", hearing of Thursday 27 May 2021.

Appendix 4: Examples of organisations offering health data services

1. SNDS

Managed by the French National Health Insurance Fund for Salaried Workers (CNAMTS), and created in 2016, the SNDS[171,172] is a pseudonymised medico-administrative data warehouse covering the entire French population and covering all care submitted for reimbursement. It is used to link data from different databases. It brings together[173]:

- health insurance (Assurance Maladie) data (based on the Assurance Maladie national inter-regime information system - Sniiram database);
- hospital data (based on the PMSI programme for the medicalisation of IT systems) historically matched with the SNIIIRAM database;
- databases on medical causes of death (database of the Epidemiology Centre on Medical Causes of Death of the French National Institute for Health and Medical Research - Inserm's CépiDc);
- data relating to disability (from the departmental centres for the disabled - MDPH - data from the national solidarity fund for autonomy - CNSA).

Access to SNDS data is highly regulated[174], and data can only be used under conditions that comply with the security standard, which aim to guarantee the confidentiality and integrity of data and the traceability of access and other processing. To this end, each patient is identified in the SNDS databases by a pseudonym, obtained by applying to the NIR (registration number for the national register of identification of natural persons) an irreversible cryptographic process called FOIN (Function for masking personal identifiers).

It should be noted that the SNDS aims to collect and provide data, but does not offer any software infrastructure or computing capacity for processing. In addition to the SNDS, there are private structures, such as HEVA[175], whose aim is to exploit SNDS data in a secure bubble, without storing them.

In a decree issued in June 2021, GIP-PDS and CNAM were designated as joint data controllers for the SNDS [176]. Under the same decree, the AgorIA Santé consortium was launched in June 2021 by Docaposte, AstraZeneca and Impact Healthcare. It received CNIL authorisation on May 23, 2022, to set up its health data warehouse with an SNDS feeder system, a first for a consortium of private players.

2. GIP-PDS

The "Health Data Platform" public interest grouping (GIP-PDS), commonly known as the *Health Data Hub* , was created by the law of 24 July 2019 on the organisation and transformation of the healthcare system. Its structure as a public interest group (GIP) brings together 56 stakeholders, the vast majority from public authorities (CNAM, CNRS,

[171]See: https://www.snds.gouv.fr/SNDS/Qu-est-ce-que-le-SNDS.

[172]See: https://documentation-snds.health-data-hub.fr/introduction/01-snds.html#les-donnees-presentes-et-absentes.

[173] See: https://assurance-maladie.ameli.fr/etudes-et-donnees/presentation-systeme-national-donnees-sante-snds.

[174]The use of SNDS data is prohibited for i) the promotion of health products, aimed at healthcare professionals or establishments, and ii) the exclusion of cover from insurance contracts or the modification of insurance contributions or premiums for an individual or group of individuals. https://www.snds.gouv.fr/SNDS/Finalites-autorisees

[175] See: https://hevaweb.com/fr/.

[176] Seehttps://www.dsih.fr/article/4763/la-creation-d-un-entrepot-de-donnees-de-sante-par-un-consortium-d-acteurs-prives-autorisee-par-la-cnil.html.

Haute Autorité de Santé, France Assos Santé, ministerial departments, etc.). The GIP-PDS implements the major strategic guidelines for the National Health Data System (SNDS) set by the French government, in particular the Ministry for Solidarity and Health. Most of its funding is public. Structured around four strategic issues: i) enhancing the value of health data assets, ii) facilitating the use of health data, iii) protecting individuals' health data, iv) innovating with all the players in the ecosystem, the GIP-PDS's service offerings aim to build genuine capacity for innovation to make France a leader in health data analysis.

3. Ouest Data Hub

In 2020, the Groupement de Coopération Sanitaire (GCS) HUGO, which brought together five University Hospitals (CHU) in the Greater West of France: Angers, Brest, Nantes, Rennes, Tours and the Institut de cancérologie de l'Ouest reached a major milestone with the launch of the "Ouest DataHub", the first hospital data platform in Europe.

The DataHub brings together anonymised data from the six member establishments to support medical research, providing an innovative way of imagining new research projects, developing personalised medicine through decision-making tools for clinicians and patients, and improving health monitoring throughout the Grand Ouest region.

4. AP-HP data warehouse

The AP-HP[177] HDW is described as Europe's first hospital data warehouse. Its aim is to bring together, standardise and structure administrative and clinical data, hospital reports, prescriptions, and the results of biological and imaging tests on more than 13 million patients treated by the university hospital centre (39 establishments), with some of the data dating back to 2012. A table updated in 2021[178] shows the research uses of the various data sets making up the warehouse. The HDW was approved by the CNIL in 2017. The AP-HP HDW is hosted in a secure[179] private cloud and relies on computing capacity adapted to the datasets and algorithms used, with 20 GPU (graphical processing unit) cards. For management purposes, access to data and business intelligence solutions is provided via the PILOTE portal, based on IBM COGNOS technology, which enables care teams, strategic departments and university hospital groups to monitor and analyse activity. The AP-HP HDW has set up a Big Data[180] platform to make the most of its data. The JUPYTER portal offers private and secure areas for processing data using standard IT languages[181]. In this sense, the AP-HP HDW is more than a data warehouse and has the characteristics of an HDP. Following the CCNE's recommendations for its governance, the AP-HP HDW has set up an Institutional Review Board, which includes patient representatives.

5. CASD

Currently relatively unknown among scientists, particularly in the public sector, the CASD (Centre d'Accès Sécurisé aux Données)[182] created by interministerial decree on 29 December 2018, is a public interest grouping bringing together the State represented by INSEE, GENES, CNRS, École polytechnique and HEC. Its main purpose is "to organise and implement secure access services for confidential data for non-profit research, study,

[177] See: https://eds.aphp.fr/nos-services/eds-donnees.
[178] See: https://eds.aphp.fr/sites/default/files/2021-09/EDS_Disponibilite_des_donnees_20210910.pdf.
[179] Proprietary structures certified as health data hosts, on AP-HP premises.
[180] See: https://eds.aphp.fr/nos-services/plateforme-outils.
[181] See: https://www.aphp.fr/connaitre-lap-hp/recherche-innovation/lentrepot-de-donnees-de-sante-de-lap-hp
[182] See: https://www.casd.eu/.

evaluation or innovation purposes, activities described as "research services", mainly public. Its mission is also to promote the technology developed to secure access to data in the private sector. In practice, the CASD is a trusted third party between producers and users of personal data, ensuring that storage and access are both secure and compliant with European regulations (GDPR).
The CASD therefore allows access to sensitive data within the meaning of the GDPR, particularly for research purposes. It is a solution that requires the implementation of specific and controlled access protocols for data and data handling.

6. Mon espace santé

Mon espace santé[183] is primarily a database for sorting information and using secure messaging, but as it also offers a catalogue of services (available since November 2022), it can be considered an HDP. The data controller and data processor are the CNAM and the Ministry of Health. The data is hosted in France by two sub-contractors: Worldline, through its subsidiary Santeos, for the shared medical record (DMP) data, and Atos for all other *Mon espace santé* data. Both of these sub-contractors are certified Health Data Hosts (HDS).

The platform was created in the public interest: "*Mon espace santé* aims to promote the role of every individual, throughout their life, in protecting and improving their health. This secure digital public service enables you to manage your health data in conjunction with health, social and medico-social sector players, thereby promoting prevention, coordination, quality and continuity of care." [184]

You can activate your *Mon espace santé* account or prevent it from being created, but to do so you need to log in to your account using the access code sent to you by CNAM. When a *Mon espace santé* account is closed, the data will be kept for ten years.

You can choose whether or not to designate healthcare professionals authorised to consult the documents, although in the event of an emergency there is a "break glass" option allowing caregivers to have access. Once a person has created a file on *Mon espace santé*, any healthcare professional consulted by the HDP user can upload a document (consultation report, analysis results, nature of examinations carried out, purchase of medication, etc.). It should be noted, however, that not all of them do so yet. Users can make their documents, medical history and advance directives visible to or hidden from healthcare professionals (except the author of the document). The confidentiality of each document can be managed, but a document submitted by a healthcare professional cannot be deleted. In the case of diagnoses and sensitive documents, access is hidden from patients until they have had a face-to-face meeting with their doctor.

Mon espace santé is not accessible for secondary uses of data, in particular research.

Finally, *Mon espace santé* is presented as the French component of the future "European Health Data Space", which is currently the subject of a proposal for a regulation by the European Parliament and the European Council.[185]

7. INCa

The French National Cancer Institute (INCa) is a public interest group created by the Public Health Act of 9 August 2004, as part of the 2003-2007 Cancer Plan, to coordinate actions in the fight against cancer. It brings together all the parties involved in the fight against

[183] See: https://www.monespacesante.fr/
[184] See: https://www.monespacesante.fr/protection-donnees-personnelles
[185] See: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_fr

cancer in France, with a twofold aim: to help reduce cancer mortality in France and improve the quality of life of people with cancer. The French government is represented by the Ministries of Health and Research. One of INCa's missions is to gather the most up-to-date information from the various data producers, and to analyse and summarise it in order to produce multidisciplinary, shared expertise on issues relating to cancer: risk factors, demographic issues, radiotherapy, genetic tests, organisational changes, the impact of technological changes, social support, etc.[186] The Cancer Data Platform (CDP) developed by INCa is a data warehouse that brings together data from different sources under the most secure conditions. Unique in Europe in terms of its quality, wealth and volume, the ambition is for it to become an essential tool for cancer research and care, to support and strengthen cancer prevention, improve care and quality of life for patients and reduce after-effects.[187]
The National Cancer Institute's (INCa) relations with the healthcare industries are strictly governed by legislative and regulatory provisions and internal rules approved by its Ethics Committee and Board of Directors. These rules are designed to guarantee the independence of INCa when, in carrying out its missions, the Institute collaborates with these industries.[188]

8. Inserm- IReSP - Aviesan

Inserm (Institut national de la santé et de la recherche médicale) has a service unit, the CépiDC (Centre d'épidémiologie sur les causes de décès), whose missions are to produce national statistics on the medical causes of death, to disseminate these statistics, and to conduct studies and research on mortality data. Its data is included in the SNDS and is openly available on the CépiDc website[189].

Inserm also participates with IReSP (Institut de recherche en santé publique) and Aviesan (Alliance nationale pour les sciences de la vie et de la santé) in the Epidemiology - France portal, which offers an online catalogue of the main French health databases that may be useful for developing public health research and expertise[190].

With the France Cohortes[191] project, Inserm will be pooling technical and human resources for eleven of its major cohorts, including Constances (Appendix 4.9).

It should be noted that Inserm has chosen to use Informatica technology for its centralised health data HDP.

9. Constances

Constances[192] is a database and HDP drawn from 220,000 volunteer participants in France. It is therefore a large-scale cohort that is open to the scientific community. French and international scientific teams wishing to use Constances for their own research can propose projects. The teams concerned mainly belong to public research bodies such as Inserm, CNRS and universities. The possibility of proposing research projects in

[186] See: https://solidarites-sante.gouv.fr/ministere/acteurs/agences-et-operateurs/article/inca-institut-national-du-cancer
[187] See : https://www.e-cancer.fr/Expertises-et-publications/La-plateforme-de-donnees-en-cancerologie
[188] See: https://www.e-cancer.fr/Institut-national-du-cancer/Deontologie-et-transparence-DPI/Le-cadre-de-la-deontologie
[189] See: https://www.cepidc.inserm.fr/
[190] See: https://epidemiologie-france.aviesan.fr/
[191] See: https://www.inserm.fr/actualite/france-cohortes-comment-perenniser-outil-recherche-exceptionnel/
[192] See: https://www.constances.fr/.

Constances is also open to research teams from industrial companies, particularly in the healthcare sector. All projects must have a public health objective and those likely to have a marketing objective are excluded..

10. UK Biobank

UK Biobank[193] is a database and HDP drawn from half a million British volunteer participants. It is therefore a large-scale cohort that constitutes a unique research resource by cross-referencing genetic information with in-depth health data. The database is regularly enriched with additional data and is accessible worldwide to accredited researchers from public and private organisations undertaking essential research into the most common and dangerous diseases.

11. Dawex

French company Dawex[194] is developing a data exchange platform enabling the distribution and co-ordination of a data ecosystem. It neither buys nor sells data, but connects companies interested in monetising and re-using data. "The platform incorporates a governance model for data exchanges, covering control, security, traceability, licensing and regulatory compliance. It can be operated in centralised, distributed or decentralised mode, and offers several business models to its participants: free, by subscription, by transaction[195]". Dawex offers a number of tools to help data providers and users assess the nature of the data being exchanged. It also offers sampling tools to combat data representativeness bias. Dawex does not currently specialise in health data, but is already used in agriculture: API-AGRO is an agricultural data-sharing centre based on Dawex technology.

12. Salus-Co-op

Salus-Co-op[196] is a Spanish citizens' cooperative, created in 2017, which offers its members the possibility of providing access to their medical information for health research projects run by non-commercial institutions, provided that they share their research data freely and without charge, and subject to donors not withdrawing their consent. With the Salus-Co-op application, data is pseudonymised, end-to-end encrypted and passes directly between users and the researchers of the projects in which they are participating[197].

13. Healthbank

In Switzerland, the Healthbank[198] health data exchange platform is a cooperative initiative that enables individuals to share their health data with whomever they wish, anonymously and securely, with the option of monetising this access. Users retain ownership of their personal health data and can decide at any time and for any reason to stop sharing their data[199]. In the basic version, opening an account is free, but you can become a member of the cooperative by buying a share (CHF 100). The platform acts as an intermediary (for a

[193] See: https://www.ukbiobank.ac.uk/.
[194] See: https://www.dawex.com/.
[195]See: https://www.dawex.com/fr/data-exchange-platform/.
[196] See: https://www.saluscoop.org/
[197] *La Croix*, (Débat), *Faut-il partager ses données au nom de l'intérêt général ? Des risques de pratique discriminatoires,* 28 June 2022.
[198] See: https://www.healthbank.coop/
[199] See: https://www.healthbank.coop/#how-it-works.

fee) between researchers and platform members. It anonymises the data, receives payment from the researchers and pays the members selling access to their data.

14. Doctolib

Doctolib[200] is a French company that began distributing an appointment management application for healthcare professionals in France in 2013, followed by Italy and Germany, as well as an online appointment booking service for patients. Doctolib has gradually developed a patient and data management service for doctors. However, in 2021, several doctors' unions lodged an appeal with the Council of State over Doctolib's use of Amazon Web Services (AWS) hosting services. However, the Council of State responded by stating that « "Doctolib has [...] set up a security mechanism for the data hosted by AWS Sarl based on a trusted third party located in France to prevent the data from being read by third parties", thereby validating the compliance of the hosting system set up by Doctolib..[201]

[200] See: https://www.doctolib.fr/.

[201] Council of State, court decision of 12 March 2021 [https://www.conseil-etat.fr/actualites/le-juge-des-referes-ne-suspend-pas-le-partenariat-entre-le-ministere-de-la-sante-et-doctolib-pour-la-gestion-des-rendez-vous-de-vaccination-contre].

Appendix 5: Data from medical research: legal framework

The formalities and procedures concerning data from human subjects research depend on the context of the research. The CNIL distinguishes between two different types of research: in-house research (carried out on patients as part of their therapeutic and medical follow-up by the care team and for its exclusive use) and multicentre research or research involving data being made accessible to third parties (for example, for a thesis or dissertation).

In the first case, research is governed by the provisions of the GDPR and must meet the requirements of chapter IX of the French Data Protection Act. In the second case, the procedure requires greater vigilance to ensure that publication of the research does not directly or indirectly identify the individuals concerned. This approach emphasises the need for those involved to take responsibility, and requires a Data Protection Impact Assessment (DPIA) to be carried out whenever the use of such data may entail a high risk for the rights and freedoms of the individuals concerned.

Clinical research in France is currently governed by the Jardé law, which provides a framework for "human subjects research (RIPH)" involving a medical device, drug or other health product. This research is classified into three categories according to the risk incurred by the participant:

- RIPH 1 (officially known as Interventional Research, RI), which concerns drug trials or trials of new implantable devices and requires specific authorisation from the French National Agency for the Safety of Medicines and Health Products (ANSM); this research involves an intervention not usually performed on human subjects;
- RIPH 2 (Interventional Research with Minimal Risks and Constraints, RIRCM) involves an intervention on a person, the list of which is set by order of the Minister for Health;
- RIPH 3 (Non-Interventional Research, RNI) does not involve any risk, as the procedures are carried out in the usual way, although a decree issued by the Minister for Health defines the procedures authorised for this category of research, which can lead to some confusion.

The Personal Protection Committees (CPP) coordinated by the National Commission for Human Subjects Research (CNRIPH)[202] are responsible for "issuing a prior opinion on the conditions for the validity of any human subjects research (trial or experiment), with regard to the criteria defined by Article L. 1123-7 of the French Public Health Code. In particular, the committees ensure that participants in human subjects research are protected (prior information, consent, exclusion period, reflection period, etc.), that the research is relevant and that the risk/benefit ratio is satisfactory"[203]..

However, there are many areas of research using health data that do not fall within the scope of the RIPH: this includes research requiring the re-use of personal health data, particularly from medical records, the SNDS or cohorts. This data may be re-used if the patient is informed and does not object.

[202] See: https://solidarites-sante.gouv.fr/systeme-de-sante-et-medico-social/innovation-et-recherche/article/la-commission-nationale-des-recherches-impliquant-la-personne-humaine-cnriph.
[203] See: https://www.iledefrance.ars.sante.fr/comites-de-protection-des-personnes-cpp.

Appendix 6: Hearings

Naomi Allen, Chief Scientist, UK Biobank;
Régis Aubry, palliative care doctor, CCNE member;
Emmanuel Bacry, Scientific Director of the GIP-PDS (Health Data Hub);
Sarah Benichou, Head of Department, Promotion of Equality and Access to Rights (Defender of Rights);
Eric Bothorel, MP for Côtes d'Armor (Renaissance), rapporteur for the parliamentary mission "For a public data policy";
Erik Boucher de Crèvecoeur, expert engineer at the Commission nationale de l'informatique et des libertés (CNIL);
Éric Chenut, President of Mutualité française;
Marie Citrini, user representative at Assistance Publique des Hôpitaux de Paris (AP-HP);
Stéphanie Combes, Director of GIP-PDS;
Caroline Cormet-Fraigneau, Vice-President for development at OVHcloud;
Marc Cuggia, public health physician, university professor - hospital practitioner (PU-PH) in medical informatics and joint leader of the GIP-PDS prefiguration mission;
Annabelle Cumyn, member of the Interdisciplinary Health Informatics Research Group, Université de Sherbrooke and Chair of the CIUSSS de l'Estrie Research Ethics Committee;
Arthur Dauphin, project manager at France Assos Santé;
Jean-François Ethier, Director of the Centre interdisciplinaire de recherche en informatique de la santé at the Université de Sherbrooke;
Valérie Fontaine, Partnerships Manager at France Assos Santé;
Guy Fournier, Public Sector and Local Authorities Director at OVHcloud;
Jérémy Greene, Professor of the History of Medicine at Johns Hopkins University;
Caroline Guillot, Deputy Director of Citizen Affairs at GIP-PDS;
Hélène Guimiot-Bréaud, Head of the Health Department at the Commission nationale de l'informatique et des libertés (CNIL);
Anne Gysenbergh-Houal, Head of Academic and Industrial Research Collaborations, AP-HP Clinical Research and Innovation Delegation;
Claudine Jacob, Director of Rights Protection, Legal Affairs, Defender of Rights;
Nicolas Kanhonou, Director, Promotion of Equality and Access to Rights Department, Defender of Rights;
Benoît Labarthe, Head of the Partnerships and Innovations Department, Research and Innovation Division, Medical Affairs, Research and Territorial Strategy Unit, Nantes University Hospital;
Laurent Lafaye, co-founder of Dawex;
Philippe Latombe, MP for Vendée (MoDem) and rapporteur for the parliamentary mission "Building and protecting national and European digital sovereignty";
Karine Lefeuvre, Professor of Vulnerable Persons Law at EHESP and Vice-Chair of the CCNE;
Franck Lethimonnier, Director of the Aviesan Alliance's Thematic Valorisation Consortium;
Laura Létourneau, Ministerial Delegate for Digital Health (DNS);
Pierre Lombrail, university professor and hospital practitioner in public health at Université Paris 13, rapporteur for the Inserm Ethics Committee's "mass data in health" working group;
Jacques Lucas, Chairman of the French Digital Health Agency (ANS);
Emmanuel Meyrieux, Head of Customer Security at OVHcloud;
Catherine Morin-Desailly, Senator for Seine Maritime (Union Centriste, UCI-UC group), involved in the issue of European digital governance;
Frédéric Ossant, project manager of the Ouest Data Hub platform;

Adrien Parrot, Chairman of InterHop;
Denis Paul, project manager at OVHcloud;
Valérie Peugeot, researcher at Orange Labs and Chair of the Vecam association, commissioner in charge of health data at the CNIL;
Christelle Rebillet, Division Manager, French Accreditation Committee;
Guillaume Ruty, IT Director, OVHcloud;
Brigitte Seroussi, Projects Director at the Ministerial Delegation for Digital Health (DNS);
Catherine Simonin, board member of France Assos Santé, member of the Ligue nationale contre le cancer;
Hubert Tardieu, Chairman of the Gaia-X Board of Directors;
Fabrice Tocco, co-founder of Dawex;
Celia Zolinsky, Professor of Law at Panthéon-Sorbonne University, member of the CNPEN.

Acknowledgements:
Bastien Rance (Imagine)