Open health: how to re-use data in the context of the Covid-19 sanitary crisis in France?

Back to Healthcare and Life Sciences Committee publications

Cécile Théard-Jallu

De Gaulle Fleurance & Associés, Paris

ctheardjallu@dgfla.com

Florence Eon-Jaguin

Withlaw Avocats, Rennes

florence.eon-jaguin@withlaw-avocats.fr

Introduction

Digital practices and the related use of personal data are now at the heart of many projects in the healthcare sector, and the Covid-19 crisis has had a boosting effect on this phenomenon. A key question is how to use such data for secondary purposes for sake of public health and innovation?

In France, the Minister of Solidarity and Health and the Minister of Higher Education, Research and Innovation, have respectively called for a free and public access to all publications and data, in all disciplines, stemming from research related to the Covid-19 epidemic, in the broadest possible sense, following the FAIR (Easy to Find, Accessible, Interoperable, Reusable) logic.[1]

However, before publishing scientific results, several steps need to be taken, both technical and legal, by public or industry players running research projects. Indeed, the use of health data presupposes compliance with the rules on the protection of personal data as well as ensuring medical secrecy and the effectiveness of the right of every citizen to privacy.

Many health data-use initiatives can be legally characterised as cases of secondary use of data. This means that the processed data was collected from individuals for a purpose that was initially different, most often consisting of the provision of diagnosis or care.

The legal framework for the re-use of health data is complex and designed to seek a balance between, on the one hand, the protection of privacy and the protection of health data, and, on the other hand, the need to promote R&D. The search for this balance results in:

- the intervention of public authorities in support of public or private database initiatives, which is even stronger in the context of a health crisis**(I)**.
- a dense regulation limiting the re-use of health data, while the need for private investments call for a sufficient level of databases' protection by intellectual property, this evolving in a context of open science sustained by both French and European authorities**(II**).

In the context of the health crisis, this balance has clearly tilted in favour of research and public health.

- –A variety of secondary use cases in the context of the health crisis with the support of public authorities

The context of the health crisis made it possible to emphasise the usefulness of reusing health data, and cases of re-use were multiplied in order to organise a response to the pandemic.

This could be observed at several levels: public initiatives encouraging this re-use (1), private initiatives taking advantage of this re-use (2) and finally individuals, who have an active role in the re-use of their health data (3).

1. Public initiatives encouraging the re-use of health data

1.1 The example of the Health Data Hub

Set up by a decree of 29 November 2019, the Health Data Hub is a platform aimed at enabling the health sector to exploit the potential of artificial intelligence by making available to researchers large volumes of health data necessary for the development of artificial intelligence models. The objective is to bring together all data from French public health organisations, such as health public insurance and healthcare centres. For users, the Health Data Hub will be a one-stop shop from which they may request access to all data registered in the catalogue.

Note that this tool does not preclude a specific authorisation application with the French data protection authority (the CNIL) if the request for access is not part of a pre-existing reference methodology or simplified procedure.[2] In order to access this platform, companies will have to justify the public interest of their project in any event.

Appealing objectives have been set for the Health Data Hub:

- Collect, organise and make available to researchers the data comprised in the National Health Data System (in French, SNDS) launched in 2016;
- Inform patients, promote and facilitate the exercise of their rights and in particular their right to object to the use of their pseudonymised data from the SNDS.

However, the establishment of the Health Data Hub has raised concerns because of the hosting of data entrusted to the American giant Microsoft, which, like all American companies, is said to be possibly forced to communicate the contents of electronic communications in their possession, pursuant to the CLOUD act. Questioned on this point during a query to the French Government,[3] the Minister of Solidarity and Health replied that this American rule should not apply 'in the context of the Health Data Hub since all the data stored in the platform is de-identified and each data set is independently stored in a space dedicated to its producer and encrypted with a key to which Microsoft does not have access'.

In the context of the Covid-19 health crisis, the Health Data Hub has found itself on the front line and the issue of the retention of personal health data for research or epidemiological purposes has been at the heart of national debates. The extension of the Health Data Hub's prerogatives, as set out in the decree of 21 April 2020, was subsequently validated by the French State Council.[4] In interlocutory proceedings contesting this extension of Microsoft's choice as hosting provider, the State Council ruled that 'the decree does not constitute a serious and manifestly illegal infringement of the right to privacy and the right to data protection'. On the other hand, the State Council ordered the Health Data Hub to notify the CNIL of the methods used to pseudonymise the data, in order to allow the CNIL to verify that those measures ensure sufficient data protection. Regarding the hosting of health data by Microsoft, the State Council considered that the applicable contractual clauses ensure sufficient protection. Nevertheless, criticisms of opponents to Microsoft's choice have been heard: Cédric O, Secretary of State for Digital Affairs, announced the implementation of a call for tenders on 23 June 2020, relating to the choice of the Health Data Hub's data host.

While the Act of 11 May 2020,[5]extending the state of health emergency until 10 July, allowed the storage of certain data for epidemiological surveillance or research on the virus to be extended by three months after the end of the state of health emergency, the order of 10 July 2020[6]postponed this period up until 30 October 2020 at the latest and allowed the crossing of databases including patients' social security numbers (which is highly regulated under French law).Article30 II of the order of July has also recalled the need to limit the use of personal data at stake to public interest projects only, in fighting against the current epidemic.[7]

1.2 Beyond the Heath Data Hub, what are public warehouses made for?

Health data warehouses are predominantly public and are scarcely interoperable. For now, only four Public University Hospitals have been authorised by the CNIL to implement automated processing of personal data for the purposes of a data warehouse: the Assistance Publique - Hôpitaux de Paris (AP-HP) since 19 January 2017 (before the entry into force of the GDPR), the Nantes University Hospital since 19 July 2018, the Lille University Hospital since 5 September 2019 and the Grenoble University Hospital since 10 October 2019.

Data warehouses are created mainly to collect and make massive data available (eg data related to medical management of the patient, socio-demographic data, data from previous research, etc.), that will subsequently be reused, particularly for the purposes of studies, research and evaluations in the field of health.

These databases are established for a long period of time (usually more than ten years) and may be fed by multiple sources (health professionals, patients, pharmacy, health institutions, etc.).

Data subjects must be provided with complete, clear and legible individual information specifically related to the establishment of the warehouse. They must also be able to exercise their rights effectively (right of access, rectification, objection, portability). Two hypotheses can be taken into consideration:[8]

- If data subject's explicit consent has been obtained for the collection, recording and storage of health data in a warehouse (for secondary use), no formalities will be necessary, but the data controller shall still demonstrate compliance of its processing with the GDPR.
- If explicit consent is not obtained, establishing the warehouse is subject to a prior authorisation of the CNIL, while the purpose of the database shall be for a public interest.

Data subjects shall subsequently be informed of each research project conducted by using the data of the warehouse. Among other possibilities, Reference Methodology MR-004 provides for the possibility of arranging the information of individuals by allowing reference to an information device (such as an internet site, for example), as long as the individuals have been informed individually beforehand.

Note that those rules also apply to private data warehouses.

1.3 Financial incentives from public authorities to support the sharing of data

The promotion of the re-use of health personal data has also been reinforced by public authorities' financial incentives.

On 15 January 2020, the French General Secretary for Investment (in French, SGPI) launched a call for projects entitled 'Let's innovate to improve the health system thanks to artificial intelligence!'[9] Opened until 24 March 2020,[10] this call for projects, co-piloted by the Health Data Hub and the Big Challenge 'Improving medical diagnostics through AI',[11] aims at supporting innovative projects, particularly in the implementation of artificial intelligence algorithms that use health data to benefit the health system. In addition to the financial support set up by BPIFrance, each project selected will be subject to 'operational support from the Health Data Hub, which may include support in the process of accessing data, support for data collection and organization, the provision of computing and storage resources, and networking with other players in the ecosystem'.[12]

Finally, the outbreak of the Covid-19 pandemic also triggered calls for innovative solutions, notably from the French Ministry of the Armed Forces through the Defense Innovation Agency (in French, AID). Launched on 19 March 2020, the call for projects focused on 'the search for innovative solutions, whether technological, organizational, managerial or for the adaptation of industrial processes, which could be directly mobilized in order to protect the population, support the care of patients, test the population, monitor the evolution of the patient at the individual level and the evolution of the pandemic, or help limit the constraints during the crisis period'.[13] Closed on 17 April 2020, the call for projects included a ten million euro budget, aimed at financing one or more projects.

Similar initiatives have also been put in place at the European level. First, in April 2020, the European launched the Covid-19 Data Portal to bring together relevant datasets for sharing and analysis in an effort to accelerate coronavirus research. As part of the first EU 'ERA vs Corona Action Plan' and its ten supported measures including the fostering of research infrastructures and databases, this portal enables researchers to upload, access and analyse Covid-19-related reference data and specialist datasets as part of the wider European Covid-19 Data Platform.[14]

The Europe Union also financially supports specific projects such as MOOD(Monitoring Outbreak events for Disease surveillance in a data science context).[15]MOOD aims at collecting mega-data in order to streamline the surveillance of public health threats. The MOOD project, funded by the European Union, builds on the exploitation of data and the analysis of mega-data to enhance the usefulness of EBS.[16]

2. Private initiatives

Within the private sphere, connected tools reusing personal health data have developed considerably. The use of health data within the framework of connected devices more particularly aims at facilitating patients' returning home more quickly and safely, for example:

- Sensors associated with an application that allows the patient's blood sugar level to be checked at any time: fixed on the insulin pen, it sends essential information such as blood sugar level, frequency of insulin doses injected, recording times, etc. The user is instantly provided with a self-monitoring logbook in graphical form, and the data can be transmitted to the healthcare professional, who can then remotely monitor the patient's condition.
- Patch equipped with sensors that continuously collects information from the encephalogram and transmits it to an application: this data constitutes the user's symptom log and informs him/her of an impending epileptic seizure. The application also benefits from a geolocation system, which allows the patient's family and friends, or the healthcare professional, to be notified directly in the event of a seizure.
- Home assistance robots for the elderly: made up of numerous sensors and a facial recognition system for emotions, the robot can detect an attack, call the healthcare professional or an assistance platform.
- Connected pill dispensers which, once the dosage has been entered, make it possible to monitor the patient's taking of the treatment, in particular by warning them when they have to take it.

Beyond these initial uses, private players marketing these tools generally aim at further using (and qualifying) this data to continue feeding and developing their artificial intelligence-based tools or sharing it with other players to create databases to support collaborative research projects.

In that respect, the Covid-19 pandemic has allowed private initiatives to flourish. For instance, platforms were set up to identify needs of healthcare establishments and competent volunteers and optimize flows management (eg, in Brittany, the medGO tool helping the Regional Health Agency organise the health reserve staff).

Initiated by a healthcare coalition of experts, including Fabrice Denis, a physician who fathered the first digital app reimbursed by the French social security scheme, Kelindi, the Pasteur Research Institute and Allianz, the 'maladiecoronavirus.fr' online testing tool was set up in March 2020 to help decrease the number of patients admitted to emergency rooms. In parallel, the 'Health Innovation - Health Crisis' coalition, supported by AP-HP and BPIFrance, was formed by France Biotech, France Digitale, MedTech in France and AstraZeneca, to help relieve congestion in the healthcare system and enable patients suffering from chronic diseases to continue to be treated, by developing and implementing innovative solutions in the field of health, based on needs identified and reported by healthcare structures, healthcare professionals and patient associations (www.coalitioncovid.org).

3. Individual initiatives

Individuals may also decide to contribute to fighting the pandemic, or more globally R&D in healthcare, thanks to their personal data.

The Embleema project, an online platform of patients' data contributing to feed clinical trials, is among the patient-driven initiatives best illustrating that phenomenon.[17]

During his lifetime, the individual has some control over the fate of his data. Indeed, although Article 6 of the French Data Protection Act (Article 9.1 of the GDPR) establishes the principle of the prohibition of processing health data, exceptions are provided for. Among other exceptions,

Article9,2,e) of the GDPR specifies that this prohibition does not apply 'to processing related to personal data which are manifestly made public by the data subject'.

In other words, the processing of health data, previously made public by the data subject, will be authorised (this will not, however, release the data controller from its obligations under the GDPR and the French Data Protection Act, including having to demonstrate a solid legal base for the re-use of the data).

The data subject may also consider the fate of his data after death. Under Article 85 of the French Data Protection Act, 'any person may define guidelines relating to the storage, deletion and communication of his personal data after his death'.

A distinction must be made between general directives, which cover all personal data related to the data subject and which may be registered with a trusted digital third party certified by the CNIL, and specific directives concerning the data processing mentioned by these directives and which are registered directly with the data controllers concerned.

The data subject also has the possibility to designate a person responsible for the execution of these directives. Failing that, the heirs[18] will be entitled to take note of the directives on the death of their author and request for their implementation on the part of data controllers concerned.

Thus, the fate of post-mortem personal data is considered from a similar perspective to that of the person's advance directives over his or her body, by virtue of the principle of the unavailability of the human body. Indeed, in the same way as for personal data, any adult person has the possibility to issue advance directives relating to the end of his/her life, concerning the continuation, limitation or stop/refusal of treatments or medical acts[19]. Would this parallel make it possible to consider that personal health data can be assimilated to the human body, and therefore unavailable?

Regardless of the public or private sponsor of the health data reuse project, and associated issues, a series of rules needs to be respected. In the context of the health crisis, some rules have been modified, while others have remained unchanged without Covid-19 being seized as an opportunity to update them, including to clarify or simplify them, while this would have been helpful to research project holders.

- - Legal framework for the re-use of health data: friction between the necessary protection of personal health data and the need to preserve private database-related investments, while an open science wave is developing fostered by the sanitary crisis

Re-use of health data for secondary purposes is legally framed between rules protecting personal data and more particularly health data and those protecting the interests of companies investing in research and database projects, while French and EU authorities push forward open science initiatives. The Covid-19 crisis is making these frictions even more acute.

4. Re-use and protection of personal data

In France, re-use of health data for secondary purpose is possible if in compliance with the GDPR and the French data protection law no°78-17 of 6 January 1978 as modified (called the 'Loi Informatique et Libertés' or 'LIL').

In particular, several legal bases (art. 6 GDPR) may be contemplated in the context of the re-use of health data: consent, public interest mission, legitimate interest.

With regard to consent, it is important not to confuse it (as a legal basis under the GDPR) with the consent that may be required under the legislation on biomedical research (ie for interventional studies on human beings with a certain level of risks as mentioned in art. L.1122-1-1 of the PHC).

Regarding legitimate interest, secondary use research may be allowed in accordance with the principles set forth by Art. 5(1)b), 6(4), 89 GDPR:

- Under article 5(1)(b) GDPR, where data is further processed for archiving purposes in the public interest, scientific, historical research[20]or statistical purposes, this shall *a priori* not be considered as incompatible with the initial purpose.
- Under Art. 6(4) GDPR, the data controller must perform and document an assessment of whether the purpose of the secondary use is compatible with the purpose for which the data was initially collected lawfully, considering in particular:

(a) any link between the purposes for which the personal data has been collected and the purposes of the intended further processing;

(b) the context in which the personal data has been collected, in particular regarding the relationship between data subjects and the controller;

(c) the nature of the personal data, in particular whether special categories of personal data are processed, pursuant to Article 9, or whether personal data related to criminal convictions and offences are processed, pursuant to Article 10;

(d) the possible consequences of the intended further processing for data subjects;

(e) the existence of appropriate safeguards, which may include encryption or pseudonymisation.

In its Opinion of 23 January 2019[21]the EDPB refers to the so-called presumption of compatibility provided under Art. 5(1)(b) GDPR. Where the prerequisites are met, the controller could be able, under certain conditions, to further process the data without the need for a new legal basis.

Under Art. 9(2) (j) and 89 GDPR, regarding the use of data for scientific research, among other criteria, research may be done by/for companies, provided:

- processing is necessary,
- interests of the data controller in processing substantially outweigh those of the data subject in not processing the data,
- appropriate and specific measures to safeguard the interests of the data subject are taken in accordance with the GDPR, LIL and other applicable legal provisions, especially the Public Health Code (compliance with medical secrecy, compliance with rules on biological research, etc.).

Under the LIL (art. 64 et seq. and 72 et seq.), in order to process health data (which is a type of special sensitive category of data) for research purposes under Art. 9(2)(j) GDPR and related provisions of the LIL, the processing must meet one of the following conditions:

- it falls within the exceptions to the prohibition of processing health or genetic data) provided under art. 9 (2) GDPR or LIL; or
- it is eligible to an MR issued by the CNIL (see below*), in which case the data controller shall file an online declaration of commitment to comply with the MR in question before conducting the secondary research project; or
- it obtains a prior specific authorisation from the CNIL for the secondary research project in question.

*The CNIL provides additional safeguards for processing health data for research purposes, including Methodologies of Reference, when no other derogatory basis allowing for the processing of health data is available.[22]Reference Methodologies ('Méthodologies de référence' or 'MR') of the CNIL include MRs 001, 002 or 003 on research involving human beings, or MR004 or 006 on research by industry players, not involving human beings/addressing data controllers whose research is for a public interest (without prejudice to the need to comply with GDPR and LIL requirements as a whole: DPIA, DPO, register, etc).

Finally, it is important to note that the anonymisation of health data avoids the scope of the GDPR: it is not simple, but it is possible and admitted from time to time by the CNIL under strict technical conditions (requiring that data subject's identity may not be retrieved in any manner).

Regulation is desirable, it is even a guarantee of confidence in digital health and in particular to enable citizens to accept that their data is reused for public health purposes, be that by public or private actors. However, despite the efforts of public authorities, which have become particularly visible during the health crisis, real simplifications still need to be made, one piece of legislation at a time and to improve the way they are linked.

5. Re-use and protection of business interests

Pressure is high on private organisations to follow an open health data and innovation path. In parallel, article L. 1111-8 of the French Public Health Code prohibits the direct or indirect sale of health data on an onerous basis. Failure to comply with this obligation is punishable by five years' imprisonment and a fine of 300,000 euros.[23]

In that context, the question arises of reconciling the re-use of personal health data for a public health purpose, on the one hand and the protection of business interests, on the other hand.

Indeed, private organisations running projects contemplating the re-use of health data will, de facto, make potentially significant investments to take advantage of this re-use, these investments frequently getting materialised through the creation of databases.

Article L.341-1 of the French Intellectual Property Code provides for the protection of *sui generis* rights of the producer of such database, understood as 'the person who takes the initiative and the risk of the investments' (financially, in equipment or in human resources). The producer's right to the database enables him to prohibit any extraction or reuse of all or part of the database (Article L.342-1 IPC). The idea is to allow the producer, who makes significant investments for the creation and maintenance of the database, to benefit from such investment's results.

This exclusive right brought by the producer's *sui generis* rights regime assumingly goes against the non-merchantable nature of personal health data contained in the database. This intellectual property protection also assumingly goes against the spirit of the digital single market, driven by the European Commission to organise Open Data at a European level. The challenge is more particularly acute regarding the European Open Science Cloud project, launched in 2018, which aims at providing a reliable environment for analysing, storing and reusing scientific databases. This is echoing a similar French initiative of Open Science launched in 2018 for the sharing of results triggered by research projects having benefitted from public funding.[24]

In this sense, a real challenge remains in this area, in that it will necessarily be less attractive for a company to make substantial investments without subsequently being able to make a return on these investments regarding the database setup.

Conclusion

On the field, a relevant balance needs to be found, which often takes the form of Business-to-Business or Public-to-Business or Business-to-Public agreements, pursuant to which an onerous user or R&D license is granted by the data base producer to its public or private partner, on the anonymised data portion or originally non-personal data portion of the database and the related documentation. The data base producer may also get paid for services that they provide while using their data base.

The data retention period will also be key: the longer the lawful personal data retention period is, the longer the database producer is able to feed the database with this data to ameliorate it, especially in the artificial intelligence field.

This duration must necessarily be long enough to ensure durability of intellectual property rights on databases and related investments and to allow the replay of scenarios for AI development purposes.

As a result, a particular attention must be paid to the need to put clear contractual provisions in place, allocating roles and responsibilities over medium and long periods of time during which health crises may occur. This concerns a variety of agreements throughout the life span of healthcare products and services: collaboration, consortium, services, health data hosting contract, licensing and technology transfers, AI tool development and licenses, manufacturing, promotion, insurance, etc.

In the same way, particular attention should be paid to the forecasting of possible cases of force majeure, as they have been made topical through the emergence of the Covid-19 health crisis..[25]