HEALTH DATA AND DIGITAL SOVEREIGNTY

Pierre-François LAGET, Lead Physician (Médecin Responsable) of the Département d'Information Médicale, Centre Hospitalier de Lisieux

May 2014. Article n°IV.5
Translated from French

There is no need to revisit the challenges posed by "Big Data," such as commercial as well as strategic challenges. The French Government was not incorrect to include big data in their evaluation as one of their seven strategic ambitions defined in the Commission Innovation 2030 (1).

The questions raised by these types of data are numerous and the challenges are above all political and strategic:

- The prominence of U.S. digital mega-players
- Legal issues (who does data on individuals belong to?)
- Possible violations of public freedoms, amplified by the different legal processes from one country to another. Data considered private by one may be considered public by another, and therefore commercially exploitable by the other.

An example presented by Pierre Bellanger (2, 3) summarizes the set of issues raised. It concerns the world of transportation. Indeed connecting vehicles to data networks is already a reality, even if for the time being, it only concerns utility vehicles or high-end products. This data will be used more extensively to manage traffic flows and thus optimize traffic, and bring many positive economic effects, if only to limit the time wasted in vehicles. The next step will be for it to be used by urban planning actors when it comes to making new lanes or urban redevelopment. However, there is one important constraint: the integration of this connectivity with mobile devices. Therefore, the vehicle operating system, which will be a significant part of their added value, must include in their design the possibility of broad interaction with connectivity tools available to users.

Specifically, the operating systems made by major internet players will be in an almost completely dominant position. They are becoming the de facto standard in areas like onboard systems, even if they had not initially set out to become that. We are witnessing a phenomena of self-reinforcing hegemony, disastrous for technological added value and above all for sovereignty.

One such example is applicable without limit. It raises the general question: *how do we define digital ecosystems to provide general interoperability while enabling our sovereign control of the rule of law?*

Specifically some answers have been given in an area where control of data is fundamental: that of health data. Simple solutions have been implemented, and this could perhaps inspire us in other areas.

GENERAL CHARACTERISTICS OF HEALTH DATA

Whatever their purpose, these types of data have three fundamental characteristics:

1) They have private status, recognized by the law, which punishes violations of this.
2) The private status has legitimate exceptions, also recognized by law.
3) They offer the opportunity to serve the public interest, including:

- allowing for public health studies on a large scale;
- providing elements for defining a national policy in regards to offering health care and the financing of health care.

But these characteristics are general. They can be extrapolated to multiple domains and, in the case of the example (of data transport), the parallel is striking:

- our movements are private;
- their analysis is nevertheless necessary to optimize traffic;
- some legitimate players are able to access it, as exceptions, within a framework established by the sovereign authorities.

We now set out the fundamental principles put in place in health data, limiting ourselves to the data produced by healthcare institutions, which are the most complex. We then consider extensions to this system.

FIRST LEVEL IN THE STRUCTURING OF HEALTH DATA: THE INSTITUTION AS THE PRODUCER

Historically, the first data producer was the independent physician. The idea of medical secrecy dates back to ancient history. Thereafter the system has become more complex with the implementation of the Assurance Maladie then the law of 31 July 1991 (Articles L6113-7 and L6113-8 of the Code de la Santé Publique, the French public health code), which requires health facilities to be assessed.

In the years ahead we will establish the first tools for identifying activities, and thanks to the environment put in place by the PMSI (Programme de Médicalisation des Systèmes d'Information) the intra-hospital databases will list information in structured datasets relating to:

- administrative patient information,
- the conditions treated,
- and the medical procedures performed.

The confidentiality of this information will be protected by a strict legal framework but will allow for certain tightly regulated exceptions.

SECOND LEVEL: INTER-INSTITUTIONAL EXCHANGES OF RECORDS

The law only allows for a very limited number of cases where personal information can be sent outside a facility. Specifically, the only times it is encountered in practice, is in cases of cancer registries, databases for epidemiological analysis, and close monitoring of serious diseases.

These registries are present at a regional level in a very small number of institutions (often CHU). The designated officials contact the doctors responsible for the medical information (l'Information Médicale) of the institutions in their region (5). Legal constraints weighing on these registries are clearly strong in regards to the transmission and storage of this highly sensitive data:

- by the strict rules issued by the CNIL;
- the obligation for these registries to be reassessed every 4 years by a national committee (Comité National) co-chaired by INSERM and the Institut de Veille Sanitaire (4).

In these cases, the security of data is ensured:

- by the legal framework that involves different independent bodies;
- by strict requirements, verifiable and auditable, regarding quality;
- by precisely defining responsibilities;
- by ensuring that those involved are trained to have in-depth knowledge of the issues surrounding this data, as well as about security systems and encryption;
- by the fact that these people are few.

Hence we see a good example regarding process management.

THIRD LEVEL: THE SUBMISSION OF DATA TO A NATIONAL PLATFORM AND ITS PUBLIC USE

Data security is protected by a high-performance tool that ensures the anonymization of data while ensuring its usability.

The idea, born in a large part from safety constraints imposed by the CNIL, is as follows (6):

- All information related to the stay of a patient (including the diseases dealt with, the acts performed, the characteristics of the stay, and the technical data that allows for the valuation of the stay) are coded according to national rules, and collected in a database of records, including the patient identifiers. This database is stored securely in the institution and **never leaves** (the only external people legally entitled to consult this are the Assurance Maladie supervising doctors).
- The standards for this database are set by regulations, the software vendors equipping hospitals are required to abide by them so they can transmit the data produced.
- This database is anonymized in the facility by using hash identifiers with three invariants: No. SS ["NIR"]; birth date; sex. Hashing uses a standard that is open and recognized and is, of course, irreversible to our knowledge.
- In addition, a number of precautions are taken to "obscure" some parts of the data without inhibiting its usability. Thus the dates of the beginning and end of hospitalization are replaced by the corresponding month and length of stay. Similarly the birth date is reduced to the year of birth, and zip codes of very small communes are further compacted in groups.
- This anonymized database is securely transmitted to a national platform, the ATIH (7).
- Within ATIH, it is again hashed twice using a secret system.
- This anonymized database that includes the records of the twenty million annual hospital admissions in France becomes usable:
    - to determine the financial allocations to each institution
    - and for a multitude of purposes relating to epidemiology, public health, all kinds of medico-economic studies, in short, overall national health policy.

Moreover, this tool is quite simple and requires few resources, resulting in an important outcome: even if it is impossible to trace the patient's identity, the successive hashes are reproducible and when they are applied to a set of invariants, **it becomes possible to follow the patient's path, regardless of the dates and without knowing their identity.**

IMPLICATIONS OF THIS MODEL

This model is thus built on two levels:

- A decentralized level, on a local level, concerning sensitive data;
- A centralized level, national, containing only anonymous data and without risks to privacy.

The result:

- The containment of sensitive data in multiple storage places, under the control of a small number of people whose responsibilities and the production process are strictly defined and regulated;
- Thus this containment makes a large-scale compromise of data unlikely;
- While at the same time, the data susceptible to being compromised is centralized and therefore is not interesting for an attacker because it is available!

Thus we see immediately the superiority of this system in comparison to the personal medical file (Dossier Médical Personnel - DMP) that centralizes sensitive data. Even if, as is the case, drastic precautions are taken to ensure its safety, a data compromise can be economically rewarding thus making it a tangible threat.

A GENERALIZED SCHEMA

Using the example of transport is perfectly possible:

- Information gathered by the vehicles remains onboard and cannot be used except through a process of the state;
- The information emitted by vehicles is anonymized by using a hash and invariant system (registration number, serial number, etc.);
- All relevant information on the traffic can then be used;
- Exchanges between vehicles and mobile consumer devices are managed with standards including an encryption system to prevent information retrieval by unauthorized actors.

IN ALL: THE SOVEREIGNTY OF DATA AS A STRATEGIC AXIS

It has become essential to implement a vigorous policy to protect private data because the stakes are high.

Major economic issues first.

The protection of data may well become a full economic model, and France has all the required technical competencies to deliver solutions and services to protect the privacy of user and ensure confidentiality. Because without data protection, confidence in the digital economy is not possible. Of course online businesses cannot develop unless users are able to have confidence in payment systems.

And let us not forget the Internet of Things, which may be a Trojan horse. Finally it is thought that the "quaternary" economy, or "economy of solutions" has considerable potential for creating value, putting quality and sustainability at the heart of the economic model (8). And without trust, success for the quaternary economy is not possible.

Strategic issues next.

France has every interest in positioning itself as **the land of digital rights**. If there is clear political will in this direction, that is understandable, can be implemented, and that is above all credible, it will be an attractive factor in addition to other competitive advantages such as the quality of our infrastructure.

If we manage to implement such a policy, we will have created a **market of sovereignty**, a direct counter-power in the face of what is clearly digital imperialism.

We will have an especially useful contribution to what will be one of the major conflicts in the coming years: to regain sovereignty over our personal data, without this the power of multinational digital companies could become intolerable.

1) http://www.redressement-productif.gouv.fr/commission-innovation-2030-installee

2) Pierre Bellanger, “La souveraineté numérique, Éditions Stock,” 2014, 264 p. (ISBN 978-2918866213)

3) Interview by Xavier Delaporte (“Place de la Toile,” France Culture, 12 April 2014)

4) http://www.invs.sante.fr/Espace-professionnels/Comite-national-des-registres/Liste-des-registres-qualifies-au-1er-mars-2014

5) Their roles and missions are defined by Articles R. 6113-1 to R. 6113-10 of the Code de la santé publique.

6) Successive versions of the decision of 22 February 2008 (l’arrêté du 22 Février 2008) concerning “au recueil et au traitement des données d’activité médicale et des données de facturation correspondantes.”

7) Agence Technique de l’Information sur l’Hospitalisation http://www.atih.sante.fr/

8) For a good overview on the subject see: M. Debonneuil, “Bienvenue dans l’économie des solutions,” http://www.paristechreview.com, April 2014.

Chaire Cyber-Défense et Cyber-sécurité (Chair of Cyberdefense and Cybersecurity)
Fondation Saint-Cyr, Ecole militaire, 1 place Joffre, 75007 Paris
Phone number: 01-45-55-43-56 - email: contact@chaire-cyber.fr;
SIRET N° 497 802 645 000 18

The Chair thanks its partners