8

Health Data in a Smart City

Vincent Keunen

Andaman7, Belgium
Email: vincent.keunen@andaman7.com

Abstract

Smart cities are cities where healthy living is encouraged. And to improve health, all healthcare professionals must agree to share the data they collect with patients. Interoperability is critical and we describe below where there is a conflict of interest and discuss anonymisation and pseudonymisation and the need for privacy and security. We describe the progress made in the USA and how Europe and smart cities could benefit from that experience. We dive into the specifics of health data and why a distributed approach is favoured by many.

8.1 Is Health Data Important for a Smart City?

Smart city and health data – what is the link? Well, if we go back to the fundamental roles of "city managers", they really are in charge of organising the life in the city, so that inhabitants live well together. And, clearly, healthy citizens will have a better life; so making sure they are healthy is important. Besides providing good care, this is also about empowering patients with their data so that they can receive better care and optionally contribute to research. So the "health data infrastructure" of a city is arguably as much important (or even more?) than other types of data.

But we have a conflict of interest.

8.2 The Conflict of Interest

In the health sector, there is a clear conflict of interest between the individual and the "common good".

Individuals want – and do have a right to – privacy. Without privacy, risks arise: the risk that you will not find a job if your employer is concerned with your health. The risk of paying higher premiums because you are at a higher risk, defeating the "solidarity" aspect of insurance, where everyone pays a little bit to cover the large expenses that some of us will face in case of serious health problems. The risk of reputation – some of us, with a public life may want to be discrete on conditions we are facing.

But for the common good, that is for finding new treatments, for improving our health systems (operationally and on the cost aspect), and for possibly other reasons, we need large quantities of data on whole populations. And that is in direct conflict with the needs of individuals.

8.3 Maybe Anonymisation is a Solution?

Is anonymisation a solution?

Anonymisation is the process of removing all PID (person identifying data) from a dataset. That can help avoid having access to one specific person's data. But, unfortunately, this is not sufficient. For example, it is sometimes needed to link data for the same person from various sources. For example, to have a long-term (longitudinal) view on the person's health – what condition they developed, when, and what is the link with their life habits, food and drugs consumption, etc. Full anonymisation is not a solution in these situations.

Pseudonymisation can help, however. What is this? It is the process of assigning a unique identifier to a given person who does not allow the viewer of the data to go back and find the identity of the person. Cryptographic techniques do exist to create a pseudonym from the identity of the person without a reverse process being reasonably available (or at least, it would take a huge amount of time with current computing resources to calculate the original PID).

It is still a risky process because some researchers have shown that, even with anonymised/pseudonymised data, under some circumstances and with additional identified data cross-linking, there are sometimes ways to re-identify anonymous data. See "Data re-identification" on Wikipedia for more details.[1]

Clearly, technological approaches to protecting citizens are only part of the solution. We need strong – and modern – laws to complement technology. It must be a big risk, with financial penalties, for companies and other institutions to use data in ways that could hurt citizens. The challenge here, as always, is that technology advances at a much faster pace than law does.

[1] https://en.wikipedia.org/wiki/Data_re-identification

GDPR is an excellent step in the right direction. And, surprisingly, many non-EU states are moving in the same direction as what EU has proposed (see the California Consumer Privacy Act,[2] several other US states privacy acts, Brazil's LGPD,[3] and many others).

8.4 Health of Citizens and Health of the City

So if a smart city needs to be a healthy city, many aspects must be considered. The environment, infrastructure, and ecological approaches to managing the city are important and are usually a focus of city managers.

But when it comes to the health of citizens, the whole healthcare ecosystem must be considered and that is sometimes less of a concern for cities. By "healthcare ecosystem", we mean hospitals and clinics – with their specialists, labs, general practitioners, nurses at home, physiotherapists, midwives, and all healthcare professionals – and maybe also caregivers, whether they are part of the family or paid professionals.

All of these players must have a role in the city, and appropriate financing mechanisms. Sometimes, cities may help incentivise some roles if they are missing or underrepresented.

Medical research is also a focus that some cities or regions could and should consider. One of the goals of medical research is to find new treatments that are effective and safe. The medical research field is already well developed in many regions of the world with pharmaceutical companies, medical device manufacturers, and bio-techs as sponsors of such research.

But cities have a role to play because some diseases are specific to some areas (malaria is only present in some regions of the world) or some ethnic groups. And it is recognised that, today, we need more diversity in medical research to improve treatments for some groups or areas.

Cities could contribute to medical research by helping the sector recruit patients and promote data interoperability in its own ecosystem. Interoperability is also critical to good care and control of costs.

8.5 Health Data Interoperability

Health data interoperability is still a very big problem in Europe today, as is in many areas of the world. However, some countries do progress significantly, like the USA (see below).

[2] https://en.wikipedia.org/wiki/California_Consumer_Privacy_Act

[3] https://en.wikipedia.org/wiki/General_Personal_Data_Protection_Law

8.5.1 Why is it hard?

Why is interoperability in health so problematic while, for example, financial institutions have solved the "financial data interoperability" problem for a long time? There are good and bad reasons for this.

Amongst the good reasons are the above-mentioned risks for privacy, for the right to be forgotten and other important elements for the citizen, for the individual. Security is a direct consequence of this: health data security is crucial to respect the privacy of patients.

Another good reason is that health data is extremely diverse. We talk about health data – versus medical data – because we need to manage – in addition to medical data from a doctor, hospital, or any healthcare practitioner (HCP) – various types of wellness, lifestyle, activity, nutrition, sleep, genetic, occupational, and medical research data (refer to Figure 8.1). Some use the more scientific term "omics" for this.[4]

Health data is very diverse and the needs vary (identified vs. anonymised data for example).

8.5.2 Unstructured data

Unstructured data includes text documents, medical images, pictures of text documents, etc., in many formats (for text documents: TXT, CSV, PDF, RTF, HTML, Word, XML (some), HL7 CDA (most), KMEHR, etc.; for images: jpg, png, gif, bmp, svg, XML, HL7 CDA, etc.).

Text documents, as such, can currently mainly be used to present the information to human readers (and transmit it). Through advanced services, that unstructured data could be converted to structured, codified data via NLP/ML (natural language processing/machine learning) systems.

8.5.3 Structured data

Structured data contain either numeric values (in many units and unit systems – imperial or metric) and codified values. There are also several codification systems: oftentimes custom but also more standard codes like LOINC, SNOMED CT, ICD, etc.[5]

---

[4] https://en.wikipedia.org/wiki/Omics

[5] LOINC: https://loinc.org and https://en.wikipedia.org/wiki/LOINC
SNOMED CT: https://www.snomed.org/ https://en.wikipedia.org/wiki/Systematized_Nomenclature_of_Medicine
ICD: https://www.who.int/classifications/classification-of-diseases https://en.wikipedia.org/wiki/International_Classification_of_Diseases

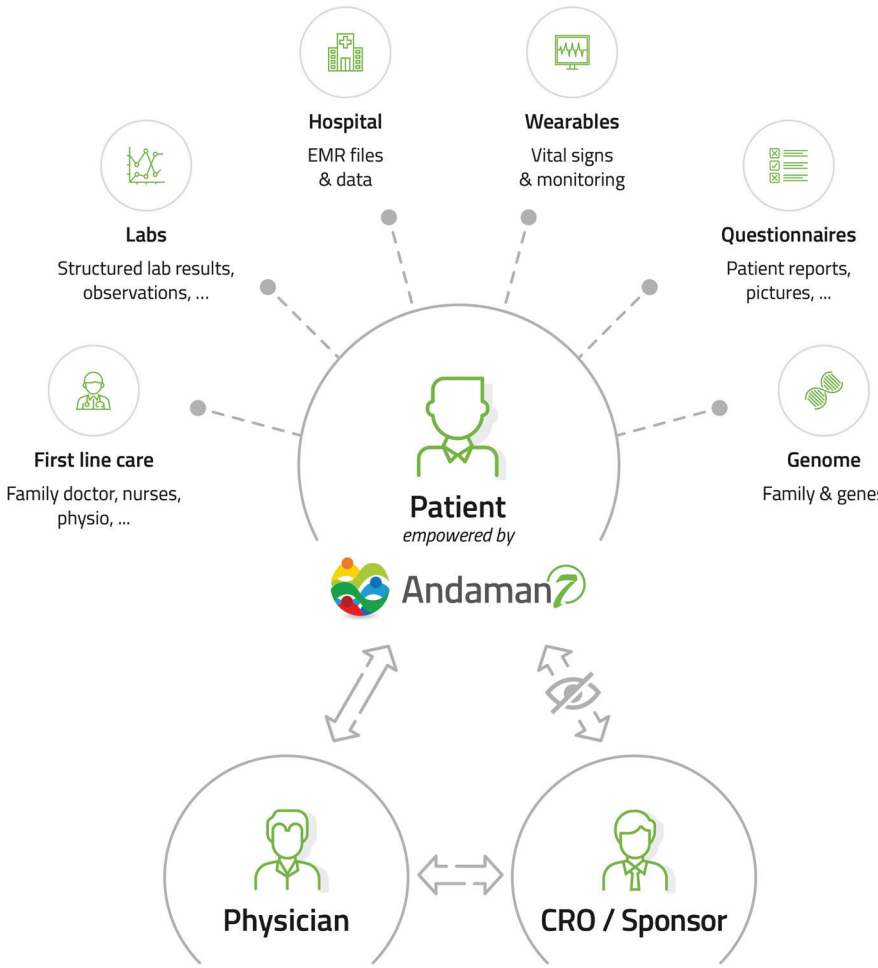

Figure 8.1 Variety of health data.

Structured data also include:

- numerous parameters coming from a growing set of wearables and connected devices, both consumer and medical grade;
- structured data resulting from rich questionnaires, trials, quality of life (QoL), ICHOM (International Consortium for Health Outcomes

Measurement), PRO (patient reported outcomes), RWE (Real World Evidence), etc.[6]

- data from Apple HealthKit on iOS devices and from Google Fit on Android devices.

But there are also less good reasons for the lack of interoperability like medical software lock-in by vendors, fears of transparency of some hospitals and healthcare professionals (revealing inefficiencies, excessive costs, or medical errors), and a simple resistance to change. But none of these serve the patients' interests and should be fought against.

8.5.4 Is the situation different in the USA?

The USA is a good example of significant progress on health data interoperability. A few years back, under the "meaningful use" initiative, all software vendors and users of these software tools were obliged to offer "standard access APIs" which means standard ways of accessing the data by external parties (other companies, patient's representatives, other institutions, etc.). Modern and mostly well-accepted standards like FHIR, LOINC, and SNOMED were chosen to improve interoperability.[7] This took a few years to accomplish.

Unfortunately, after these efforts, many hospitals still resisted sharing their data with patients. In early 2021, several patient associations made it clear to the US government that there were many impediments to them for accessing their data and the government passed new laws to impose financial penalties to these organisations (see "Denying Patients Access to Health Records/Exceeding Timescale for Providing Access" at https://www.hipaajournal.com/common-hipaa-violations). Now, the situation is finally getting better for patients in the United States and hopefully Europe and smart cities may draw lessons from their experience.

---

[6] QoL: https://en.wikipedia.org/wiki/Quality_of_life
ICHOM: https://www.ichom.org/
PRO: https://en.wikipedia.org/wiki/Patient-reported_outcome; RWE: https://en.wikipedia.org/wiki/Real_world_evidence and https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence

[7] FHIR: https://hl7.org/fhir/ and https://en.wikipedia.org/wiki/Fast_Healthcare_Interoperability_Resources
LOINC: https://loinc.org and https://en.wikipedia.org/wiki/LOINC
SNOMED CT: https://www.snomed.org/ and https://en.wikipedia.org/wiki/Systematized_Nomenclature_of_Medicine

8.6 The InteropEHRate Project

To help improve the situation in the EU, the InteropEHRate project, described in Chapter 2, has developed a number of protocols to improve cross-border interoperability.[8]

The project enables patients to be in full control of the usage and the routes of their health data. The central instrument, being laid in "patients' hands", is the smart EHR (S-EHR), leveraging a set of new protocols for secure and cross-border exchange of health data. Andaman7 is the reference implementation for the S-EHR.

8.7 Data Ownership and the Distributed Approach

Besides interoperability, there is also a lot of discussion on "data ownership". Is data generated by a hospital or a healthcare professional their property? It is especially critical because that data is about an individual person and their health. It is only ethical that pretended data ownership does not interfere with a person's good health. Retention or not sharing data can have dramatic impacts on patients, greatly reducing the quality of care or its cost. The healthcare industry should never abuse their power in ways that could hurt patients. Patients are citizens in a fragile moment of their existence. Healthcare is not and should never be an industry like any other. Ethics must play an even more important role than in the general industry.

At the same time, we should not forget that the patient is paying the care provider – directly or indirectly via their insurance or social security. It is only natural that, as a result of this service, patients receive the information and data associated with their diagnosis and treatment.

Data ownership is not the right question, actually, and European laws for protecting patients and citizens move in that direction. Data is not a physical element. It can be easily (and with almost no cost) duplicated. Access to data can be given and access is becoming the crux of the question. It does not matter that much who "owns" a piece of data, but it is important to define "access rules". And GDPR is clear: patients should have access to their data, be able to correct them, have a right to be forgotten, etc.

So patients should have easy access to their health data – whoever is the "owner". And it should be accessible in a FAIR format. FAIR data are data which meet principles of findability, accessibility, interoperability, and

[8] https://www.interopehrate.eu
http://www.andaman7.com

reusability.[9] That means that patients should have access to the highest quality data and in a structured and/or codified format if they are available at the source. It is not good enough to give paper copies to patients. And it is not good enough to give them a PDF report from their lab results (or from any other data type). Both the PDF (for readability, comments, validation information, suggestions, etc.) and the underlying structured data must be provided (for example, in the FHIR format, with an LOINC codification). The structured data can then be processed on the PHR of patients. For example, with structured data, a patient can follow their PSA values over the years to control the evolution and the risk of prostate cancer.

Where should that data be stored? In the hospital? On national servers? Here, again, there have been many discussions over the years. A small number of countries (Denmark, Estonia) have implemented national systems with a central storage of all health data. This has the benefit of having a well-organised system, with advanced capabilities to process data for care and cost optimisation but also for research (also known as "secondary use of data").

However, many countries resist centralisation of health data, for fears of malevolent use of the data, and fears of too much information being available to the government – and the hackers that would succeed in breaking into those centralised systems.

Taking this into account, some stakeholders have proposed a distributed approach with no master/slave architecture.[10] This is a bit more challenging technically, but in the age of artificial intelligence and machine learning, it is actually very accessible. In the distributed architecture, data is copied to every location it is needed. The exchange of data is based on a peer-to-peer approach. With advanced traceability techniques, it is actually possible to have copies of data in various locations without redundancy or data conflicts. Data can be added to the EHR (electronic health record) of a given patient at any location, by anyone (who has the rights to do it), at any time. Data will then flow to the other locations it is needed.

This distributed approach is pretty novel, but it has been shown to be very effective. Patients using the PHR (personal health record) can collect their data from many sources, manually add information to it, and then share that data back to persons that are in their circle of trust. The traceability of the

[9] See https://en.wikipedia.org/wiki/FAIR_data for more details

[10] (see Andaman7 at http://www.andaman7.com

data is then used by healthcare professionals to give varying degrees of credit to the information at hand, depending on the source of the data.

The distributed approach is the only one to be future proof. Each stakeholder needs to stop thinking that they are the centre of the world and have their IT systems built in this way. We now live in ecosystems where several players can contribute to the health of citizens.