Between Data Governance and Artificial Intelligence: What Place for the Pursuit of the General Interest?

Proceedings of the Closing Conference of the Chair for Social Justice and AI

Under the scientific direction of
Anne-Sophie Hulin

Fondation Abeona
Championing Responsible AI

ENS | PSL

obvia

UNIVERSITÉ LAVAL

Authors:

Anita Burgun: MD, PhD, Professor of Medical Informatics at Université de Paris. Chair of the Department of Medical Informatics at Georges Pompidou and Necker Hospitals (AP-HP).

Stéphanie Combes: Director, Health Data Hub.

Nathalie de Grove-Valdeyron: Professor of Public Law, Toulouse Capitole University (IRDEIC), Jean Monnet Chair (2023-2026).

Caroline Guillot: Director of citizen Engagement, Health Data Hub.

Anne-Sophie Hulin: Holder of the Social Justice and AI Chair, professor at the Faculty of Law of Université de Sherbrooke and co-head of the Law, cyberjustice and cybersecurity axis of Obvia.

Jacques Priol: Chairman of CIVITEO and of the Data Publica Observatory, associate expert with the Global Partnership on Artificial Intelligence.

Jeanne Solofrizzo: Consultant with CIVITEO and master's student in digital public law at the University of Paris I Panthéon-Sorbonne.

Grimaud Valat: Associate Lawyer, DTMV Avocats and Senior Advisor, Human Technology Foundation.

This publication was produced as part of the work of the Social Justice and Artificial Intelligence Chair, financially supported by the Abeona Foundation, Obvia, Université Laval and the École normale supérieure de Paris (ENS-PSL). This work also received financial support from the Fonds de recherche du Québec.

Produced with the financial support of the Fonds de recherche du Québec

ISBN: 978-2-925138-54-9
DOI: 10.61737/NEAG1390

To cite this document: Hulin, AS (dir), (2024). *Between Data Governance and Artificial Intelligence: What Place for the Pursuit of the General Interest?* Obvia. https://doi.org/10.61737/NEAG1390

Table of contents

Foreword

Anne-Sophie Hulin

Holder of the Social Justice and AI Chair, professor at the Faculty of Law of Université de Sherbrooke and co-head of the Law, cyberjustice and cybersecurity axis of Obvia.

Excitement, stupor and trembling, passion and (un) reason...

Here are a few words that, to my mind, capture the tumultuous feelings generated by recent advances in artificial intelligence and the societal issues they raise. Alongside a lively enthusiasm for AI's potential to transform our lives and our industries, its reception is as much one of amazement at the speed of progress as it is one of concern that all innovation comes at a social price. The explosive growth of AI and the use of big data has indeed exacerbated and revealed complex societal issues, including privacy, data security, the transformation of industrial and creative activities, respect for individual autonomy as well as implications for employment and social inequalities, etc. These issues raise ethical, legal and technical dilemmas about how technology should be developed, deployed and regulated so as to guarantee a satisfactory level of compatibility between collective well-being and the protection of individual rights and freedoms. Striking a balance between technological innovation and respect for ethical and legal principles constitutes a serious challenge for modern societies.

A postmodernity of data governance and AI?

From this point of view, the last few years have been a pivotal time in terms of data and AI governance, given the growing collective awareness of the ethical, social and economic implications of these technologies. The need to regulate technological innovation has thus become clearer than ever, particularly with the large-scale deployment of generative AI among the general public and the difficulty in mastering general purpose AI. As a result, all over the world and on different levels (national, regional or international), there has been an abundance of reflection and initiatives on the subject of regulation. Among these, the legal offensive of the European Union is especially noteworthy. Through the construction of a rich, innovative and strict legal arsenal, it has equipped itself with the first regulatory framework for data governance and artificial intelligence. Indeed, alongside the European regulations on data governance[1] and on data,[2] the European Union is in the process of adopting the first regulation to determine the conditions under which artificial intelligence systems can be marketed. Subject to the final stages of the European legislative process, entry into force is scheduled for 2026.[3] Canada has also opted for the legislative approach, following the lead of the European Union. The *Artificial Intelligence and Data Act* (AIDA), which has been criticized for its form and content, is still being drafted (PL c-27)[4].

1 Regulation (EU) 2022/868 of the European Parliament and of the Council of May 30, 2022 on European data governance and amending Regulation (EU) 2018/1724.

2 Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 concerning harmonized rules on the fairness of access to and use of data and amending, Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Regulation).

3 "Artificial Intelligence Act: MEPs adopt landmark law" "Artificial intelligence: MEPs pass historic legislation", *European Parliament* (March 13, 2024), online: <https://www.europarl.europa.eu/news/fr/press-room/20240308IPR19015/intelligence-artificielle-les-deputes-adoptent-une-legislation-historique >.

4 *An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts*, Bill C-27 (second reading and reference to committee - April 24, 2023), 1e sess., 44e legis. (Can.).

If adopted, AIDA would be the first law in Canada to impose common requirements on the design, development and deployment of intelligence systems. Compared with the European regulation, AIDA has a more limited scope of application, and it is not clear what its real impact will be, pending the publication of the regulations that will define the bulk of its application. Other jurisdictions have also adopted legal texts, but not to the same extent (e.g., China, the United States). Some, on the contrary, are refusing to do so in order to preserve innovation without ignoring the potential abuses of AI (e.g., the United Kingdom)[5]. Regulatory agility is thus being prioritized. As AI and data governance models proliferate, a growing trend is emerging towards developing a framework that balances innovation with the protection of individual rights and freedoms. This framework aims to ensure the deployment and use of artificial intelligence benefits society as a whole.

The general interest as a lever for regulating the use of data and the development of AI

This collection is set within a contemporary context where defining the directions and objectives for the governance framework of data and artificial intelligence is of paramount importance. The contributions gathered here illustrate some of the discussions that took place at the symposium organized by the Social Justice and Artificial Intelligence Chair, held on February 8, 2024, at the École Normale Supérieure in Paris. Bringing together experts from a variety of fields, the symposium aimed to explore some of the points of intersection between data, AI, and the general interest. This leads to the following questions: to what extent can the pursuit of the general interest legitimize digital innovation? How can it influence its regulation? What social perspectives and/or ways of tempering current practices does it offer from the point of view of data use and AI development? In other words: *between data governance and artificial intelligence: what is the place of the pursuit of the general interest?*

The afternoon's debates focused on three main themes, all of which reflect concerns of general interest. Indeed, although the determination of the general interest is fraught with complexity, particularly in a context of comparative law, it remains possible to identify concrete manifestations of it.

In doing so, the first panel focused on the opportunities and challenges of data altruism. This concept refers to the idea of voluntarily sharing personal or corporate data for purposes that are beneficial to society, without seeking personal profit. This sharing can be motivated by reasons which are ethical, humanitarian, or pertaining to scientific research, where data can be used to improve services, develop new technologies, or assist in public health, crisis management, or sustainable development studies. Since its introduction into European Union law with the European Data Governance Act, data altruism has come under scrutiny. Not only has it established itself as an alternative mode of access to individuals' data, but it is also one of the levers of the European data strategy established to support technological innovation.

Data altruism suggests a reorientation of practices and policies around data collection, analysis, and sharing, favouring an approach that values the contribution of data to the general interest, without neglecting the protection of individual rights. In this context, data altruism becomes a catalyst for rethinking information governance in the digital age and finding levers to ensure transparency, accountability, and citizen participation. Indeed, data altruism contributes to the digital empowerment of individuals who, through their data donations, take control of the secondary use of their data.

5 *A pro-innovation approach to AI regulation*, Policy paper, Presented to Parliament by the Secretary of State for Science, Innovation and Technology by Command of His Majesty on 29 March 2023, 2023, online: < https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper >.

However, this alternative mode of data circulation also raises important questions about privacy and data security. It requires clear regulatory frameworks more than ever to protect individuals' rights while also fostering an environment of trust and inclusion. Without these key elements, initiatives aimed at promoting altruistic data sharing do not benefit from a sufficiently incentivizing context for individuals and present a high risk of failure. Its feasibility is also subject to doubts. How can one share their data when they do not possess or control it? Data altruism thus highlights the importance of making the right to portability fully effective. Otherwise, it will remain an unfulfilled ambition.

The second panel focused on the ambivalent link between AI development and environmental protection. AI, with its ability to process vast quantities of data and rapidly perform complex tasks, does indeed present considerable opportunities for innovation and the improvement of services in many sectors. However, its rapid expansion also raises important questions about energy consumption and resource use, as well as ethical and social implications. While the role of AI in understanding and managing environmental issues is usually in the foreground, the environmental footprint of its development raises the question as to what extent AI should and can continue to develop within societies.

The discussions in this panel have commonly highlighted the importance of maintaining a critical, if not reasoned, approach to the development and use of AI. Is the use of AI by organizations always necessary? It seems imperative to consider specific contexts and particular needs before deciding on its systematic adoption. For, as promising as AI may be from a social, economic, and environmental perspective, it appears crucial, given the increasing data and energy demands of AI systems, to conduct a holistic evaluation of AI solutions. This evaluation should not be limited to their technical efficiency or economic and social potential but should pay greater attention to their combined social, economic, and environmental impacts. Such an approach involves considering how AI technologies can be designed and deployed in ways that promote collective well-being, both present and future, while adhering to the principles of social justice. Such alignment is even more relevant given that AI technologies are not neutral: they reflect the values of the societies that develop and use them. In addition to governance frameworks that integrate these considerations, the development of methodologies to support organizations wishing to opt for artificial intelligence also seems to be an avenue that should not be underestimated in pursuit of guaranteeing the responsible use of AI.

Finally, the third and last panel focused on healthcare, which represents a central meeting point between data governance, AI, and the general interest. Healthcare represents fertile ground for AI integration, marking a crucial convergence between data governance and the general interest. Both the secondary use of data and the implementation of AI in healthcare offer unique opportunities in terms of improved diagnosis and personalized treatment, efficient management of healthcare resources, and so on. This explains why the European Union has made the creation of the European Health Data Space a priority project, aiming to combine public initiatives with data altruism.

However, the secondary use of health data raises difficult ethical and legal questions concerning confidentiality, data security, equitable access to technology, the patient-practitioner relationship, as well as major operational difficulties (access to relevant data, security, data interoperability). These multiple issues raise the challenge of striking a socially acceptable balance between innovation and the protection of individuals, in a context where the links between data, AI, and healthcare are continually tightening. This balance is even more difficult to achieve because health data are indeed sensitive but also marked by an intrinsic hybridity. While they are at the heart of a person's privacy and health, they are fundamental from the perspective of collective health. Thus, the expectations for protection are inversely proportional to the need for their secondary use for research purposes. It is therefore necessary to find a middle ground between individual confidentiality and collective benefit, in order to enable significant progress while respecting individuals' rights.

Together, the discussions highlighted the crucial importance of ethical, transparent, inclusive, and sustainable data governance, as well as the transformative potential of AI in addressing major societal challenges such as improving health and protecting the environment. All in all, each of the contributions subtly examine the various tensions surrounding the relationship we have with our data, underline the importance of collective responsibility in data governance and the development of AI, and illustrate how digital technology is renewing citizen involvement.

The Social Justice and Artificial Intelligence Chair

In concluding this foreword, I would like to express my deep gratitude to every speaker at this symposium, and especially to those who agreed to publish its contents. Their generosity in sharing their perspectives and research have profoundly enriched our understanding of the ethical, social, and environmental challenges associated with artificial intelligence from a social justice perspective. These exchanges have not only laid the foundations for future research, but have also fuelled an essential dialogue on the ways in which artificial intelligence can be directed towards promoting social justice. There is no doubt that the publication of these conference proceedings is a valuable resource for anyone interested in the intersections between data governance, artificial intelligence, and the common good. I'm therefore delighted that, in addition to the conference recordings freely available online, an open-access publication, in both French and English, has been produced so as to widely share the valuable contributions of this symposium.

Finally, my concluding words are intended to highlight the exceptional opportunity that the Social Justice and AI Chair represented in my (still young) research career. Each year, the Social Justice and Artificial Intelligence Chair welcomes a researcher to explore the intersections between AI and social justice. Since its founding, the Chair has had the honour of hosting researchers such as Kate Crawford and Karine Gentelet, and I had the immense privilege of being the third holder for the year 2022–2023. I am extremely grateful to the Chair's partners—the Abeona Foundation, ENS PSL Paris, Obvia, and Université Laval—for their unwavering support and trust. Their commitment to research for AI that respects the rights and interests of both individuals and the community has been a continuous source of inspiration and has contributed to enriching my reflection and research. In this regard, I wish to express my gratitude to Anne Bouverot, Tanya Perelmuter, Thierry Poibeau, and Lyse Langlois for their valuable guidance in conducting my work.

I would also like to extend my gratitude to the entire operational team at Obvia, especially Guillaume Macaux, for their support in the production and broad dissemination of my work.

Additionally, I wish to thank the team of researchers at the Chair (Alexia Argiolas, Émilie Bélanger, Émilie Guiraud, and Samantha Dorinet). I commend their responsiveness in research given the frenetic legal developments of recent months in this area.

The organization of the conference on February 8, 2024, and the realization of this book also benefited from the invaluable assistance of Samantha Dorinet. The editing of the book also benefited from the support of Daniel Polillo for the English language revision.

Check the event page for more information

The inaccessibility of abundance: Tantalus' torment

Grimaud Valat

Associate Lawyer, DTMV Avocats
Senior Advisor, Human Technology Foundation

Where the abundance of data collected and exploited for commercial purposes must be put in the service of the general interest through a voluntary approach by the individual.

What a torment it is to be surrounded, almost overwhelmed, by the resources you vitally need, without being able to access them...

Such was the fate reserved for Tantalus, son of Zeus punished by the gods of Olympus to spend eternity in Tartarus, standing in the middle of a river without being able to drink from it, and surrounded by fruit trees whose branches slipped away from him as soon as he tried to touch them.

Tantalus, the torment of inaccessible abundance.

Isn't this the ordeal suffered today by any project leader who is essentially focused on the general interest, and who needs data to meet the objectives he intends to achieve?

What a paradox it is to note the almost infinite abundance of data collected without difficulty for commercial purposes, and the inaccessibility of the same data for altruistic purposes!

To illustrate this, let's take the simple, concrete example of people's traffic data within cities. This data is collected, processed, and used by navigation application publishers to feed their artificial intelligence systems and to try to offer their users optimized and less congested routes.

This data exists, and is frighteningly accurate, but it is not accessible to the public institutions in charge of designing traffic plans adapted to new forms of mobility...

The lack of sharing of this data leads to the promotion of services to bypass traffic jams, where they could be avoided.

Sharing this data could simultaneously improve the quality of life of city residents and passers-by, while helping to reduce CO^2 emissions.

If we are to overcome this paradoxical situation, we need to understand its origins: the major distrust on the part of individuals and companies alike with regard to the sharing of their data.

The creation of this mistrust was inevitable considering the way people were initially confronted with the large-scale use of their data.

Let's not forget that our first contact with the large-scale use of our personal data was with social networks, private providers of "free" entertainment services.

And individuals poured out everything that constituted their private lives, including particularly sensitive information such as their political opinions or sexual orientation, to share it with their digital "friends," though shielded from others by a screen (of smoke?). Since nothing is truly free, the underlying model was obviously to grant the publishers of these services the right to use the data poured into them to create and feed the most sophisticated targeted advertising system ever created.

Personal data became the "black gold" of the 21st century, with its exploitation propelling startups in record time (a few years at most) to the ranks of companies with revenues and influence surpassing those of entire nations.

This expression, "black gold", perfectly reflects the purely mercantile vision we have of the available data and carries with it everything that is negative and unfair about the exploitation of these resources.

Yet data is not black gold and its exploitation cannot be limited by this purely economic vision, where its collection and sharing can be used to solve some of the greatest challenges facing our civilization.

Sharing data must be used to improve health, mobility, and of course, the fight against global warming.

But this can only be achieved by overcoming mistrust.

The educational effort is a decisive element here and must serve to demonstrate a concept that is far removed from what forms the basis of our usual reasoning: when it comes to data, protection enables sharing.

Unlike rival goods, which deteriorate or even disappear with use and whose protection is an obstacle to sharing, data are non-rival goods and can be exploited and shared without disappearing or deteriorating.

Data protection is not about exclusive use, but about the governance of that use.

This applies to both individuals and companies. The former want to preserve their privacy (in the equally paradoxical context of the large-scale use of social networks) and free themselves from any state "surveillance". The latter are seeking to preserve their competitive edge.

A solution must therefore be found to ensure data protection, not only through technical measures but above all through governance that safeguards the interests of all stakeholders.

Data altruism offers a governance-based solution to ensure that data is protected in a way that allows it to be shared for the public interest.

The European Commission, through the Data Governance Act, proposes data altruism as a third way of sharing data, distinct from commercial exploitation and regulatory mandates for compulsory sharing.

The principle is based on the creation of a trusted third party: the data altruism organization.

To qualify for this title, the organization must meet a number of criteria, including being a non-profit, being legally and functionally independent of any other organization or activity, and of course, being created to serve a general interest purpose that it must declare.

The role of this organization is then to collect data from contributors and then to organize the governance of this data so that it benefits the general interest purpose which promotes, while ensuring that the contributors' rights are respected.

These conditions counter most of the causes of mistrust: data collectors are not profit-seeking, they are independent of any company or state, and they govern the data entrusted to them to serve the general interest, protecting it so that it can be better shared.

The conditions for trust are thus created and supported by transparency obligations enabling contributors to know how and by whom their data is used.

The other distinctive feature of data altruism is that it is based on a voluntary approach by contributors, an active sharing process that is neither imposed nor forced but driven by choice and a desire to contribute to a cause.

The European Commission is thus creating a regulatory framework designed to encourage the emergence of data-sharing initiatives serving the general interest in a context that fosters the emergence of trust.

If trust can be regained, it should be stressed that in addition to mistrust, there is a second major obstacle to data sharing: the control of this data by its subjects.

It's obvious enough: in reality, individuals have extremely limited control over their data, if any at all.

Data is collected by digital service providers but is not directly accessible by individuals.

But sharing implies control.

Establishing a trusted digital identity centred on the individual is probably one of the best ways to address this problem.

In addition to the individual identification and authentication services enabled by digital identity systems, they also offer the possibility of bringing together in a single secure location, which is known as a "digital wallet", the data and attributes forming what is known as an individual's digital identity in the broadest sense of the term.

The creation of such a digital wallet, such as the one currently being promoted by the European Union, would give citizens a single channel for accessing and controlling their data while benefiting from a high level of security (or at the very least, a substantial level) as well as the authentication of the data it would contain.

As one of the levers of trust is the voluntary nature of the use of technology, it is important that digital identity systems are designed for individuals and that said individuals can decide

- what they intend to store in their digital identity portfolio
- and then what they intend to share, with whom they intend to share it, and for what use.

If official identification of the individual is enabled by the primary functions of the digital identity system (understood as a digital extension of ID cards and other passports), this can authenticate the link between the data contained in the digital wallet and its holder. As a result, some of this data, and in particular certain attributes, can be shared without the full identity of the discloser having to be revealed, in strict compliance with the principles of data minimization and privacy protection. This considerably increases the possibilities offered by such a tool while ensuring maximum personal protection. For example, proof of age, whether required online or in the real world to access certain places or services, becomes possible without disclosing any information other than compliance with the age limit. Once the person has been identified and their age data has been certified, it is possible to create an anonymized age certificate.

Such a system also enables individuals who control the data contained in their digital portfolio to share it (for example, with data-altruistic organizations), having earned their trust thanks to the rules of governance put in place and the transparency reports made available to contributors.

The data altruism organization, on the other hand, knows that the data transferred from a digital identity portfolio comes from an individual whose identity has been verified and that this data is certified. It can therefore collect the data and make it available for a purpose that serves the general interest, without having to know the identity of the contributor and with a high degree of trust in the data.

This shows that when data is protected and governed from end to end, it can be shared.

Protected data can be shared, and shared data can remain protected.

It is only in this way, by understanding the issue of sharing data in the service of the general interest as a global system involving the creation of an end-to-end framework of trust, that we will be able to put an end to our Tantalian torment and finally gain access to an abundance which has been liberated in the service of the common good.

Bibliography:

Human Technology Foundation reports:

- On data altruism, online: <https://www.human-technology-foundation.org/fr-news/rapport-data-altruisme>.
- On Digital Identity, online: <https://www.human-technology-foundation.org/fr-news/rapport-sur-lidentite-numerique>.

Trusted AI and public action: choosing useful AI first

Jacques PRIOL

Chairman of CIVITEO and of the Data Publica Observatory, associate expert with the Global Partnership on Artificial Intelligence

Jeanne SOLOFRIZZO

Consultant with CIVITEO and master's student in digital public law at the University of Paris I Panthéon-Sorbonne.

The year 2023 will go down in the history of technological progress as that of the irruption of generative artificial intelligence at the heart of scientific, media, political, and societal debates. In the space of a few months, in France and elsewhere, almost everyone heard about ChatGPT, followed by other large language models. By May, it was still a period of great amazement: 71% of French people had heard of ChatGPT, but 72% considered themselves insufficiently trained to use it. Note that 68% of those who use it at work do not tell their direct superior.[1]

Now come the first set of controversies. First of all, in Europe, they concern the compatibility of these tools with the General Data Protection Regulation (GDPR). Did the training data include personal data collected outside of any lawful consent framework? More worryingly, how can the use of these generative AIs guarantee the legality of the processing of its own users' data? These questions, posed by the regulatory authorities in various European countries, will lead to some adjustments and delays in the deployment of OpenAi or Google's LLMs in Europe. One major French city (Montpellier) has even temporarily banned the use of ChatGPT by all of its public servants.

Other debates are emerging on a global scale. Learning databases contain a great deal of protected data, so the use of generative AI exposes its users to lawsuits for infringement of copyright or of trade secrets. The energy cost and carbon impact of each model, but also of each query, are gradually being revealed. More importantly, after the initial astonishment, people are becoming aware that these AIs are not infallible. Worse still, they suffer from hallucinations. This point is essential, because to understand it you have to get inside the black box. We need to explain the principles of machine learning and understand that these generative AIs calculate probabilities that a text (an image, a sound, a video...) is plausible. And thanks to the emphasis placed on these violently disruptive tools, it has become possible to explain other, less spectacular artificial intelligence systems, which have begun to be deployed in many areas of our society with little fanfare.

Local public services in France weren't expecting ChatGPT to take an interest in AI. But hardly anyone was talking about it. A few pioneering territories, often large regions and metropolises had tested algorithmic learning tools to model and predict situations and optimize public management. A case in point is the Occitanie region, which has been using AI since 2020 to analyze the training needs of jobseekers so as to rapidly propose internships and training paths tailored to the immediate needs of businesses. Further examples include Nantes and Toulouse, who are working on predicting school canteen attendance to combat food waste.

1 IPSOS survey - May 2023

In 2022 and 2023, the Data Publica Observatory[2] conducted a major survey of French local authorities on the use of data, open data, cybersecurity, and, of course, the use of artificial intelligence. A portion of these results is presented here.

This is followed by a focus on the conditions for AI deployment by French local authorities. The analysis tool presented was developed from the CIVITEO firm consultants during workshops with the Centre Val-de-Loire region, the Paris-Saclay agglomeration, and the towns of Angers, Montpellier, Montreuil, and Clichy-sous-Bois.

I. Current status of AI use by local authorities in France

The Data Publica Observatory survey was administered during the summer of 2023 to two hundred local authorities in France (regions, departments, communes or *établissements publics de cooperation intercommunale*) spread across the entire metropolitan territory.[3]

27% of local authority respondents say they have already used (or tested) AI, compared with 15% in 2022. Unsurprisingly, these are mainly large or very large local authorities. Metropolises[4] are the most numerous to have tested AI (44%). So have 36% of departments and 30% of regions. By contrast, no local authority with fewer than 10,000 inhabitants reported having tested AI. In all, by mid-2023, there were more than 50 regional AI projects in the process of being deployed, compared with 19 in 2022[5] and just 5 in 2021.[6] Priority topics include mobility flow analysis, conversational robots for users, modelling tools to optimize waste management, canteens, and public lighting. Several projects concern security, notably the use of computer vision, which has been the subject of controversy due to the adoption of a law in France in that authorizes its use to prepare the hosting of the Paris Olympic Games.

But we are about to see that this snapshot from the summer of 2023 will soon be outdated. Generative AI is the reason!

In the Observatory's survey, 30% of local authorities say they intend to test AI in 2024. The three top themes concern climate change: using AI to steer environmental policies, energy management, and water management. During 2023, the French government launched a call for projects to finance regional AIs in service of climate transition. Each project selected is likely to benefit from several million euros in subsidies, provided it can demonstrate reproducible results. The projects are numerous. They concern the management of water tables, the reduction of energy consumption by public services, the impact of urban planning rules on soil artificialization or biodiversity, the reduction of automobile mobility, the improvement of waste sorting and recycling, etc. At the same time, the large-scale dissemination of AI systems enabling image analysis in conjunction with video-protection systems is continuing, not only for security purposes but also for intelligent traffic fluidification (priority to bicycles or public transport, for example) or the detection of illegal waste dumps.

2 The Data Publica Observatory is a non-profit association founded in 2020 by the law firms CIVITEO, DATACTIVIST, INNOPUBLICA and PARME AVOCATS. The Observatory is one of OBVIA's partners in France.

3 The full survey and its data are available online: < https://enquete.data-publica.eu/rapport/Rapport.html>.

4 In France, metropolises are local authorities with special status for the management of the 21 largest urban areas.

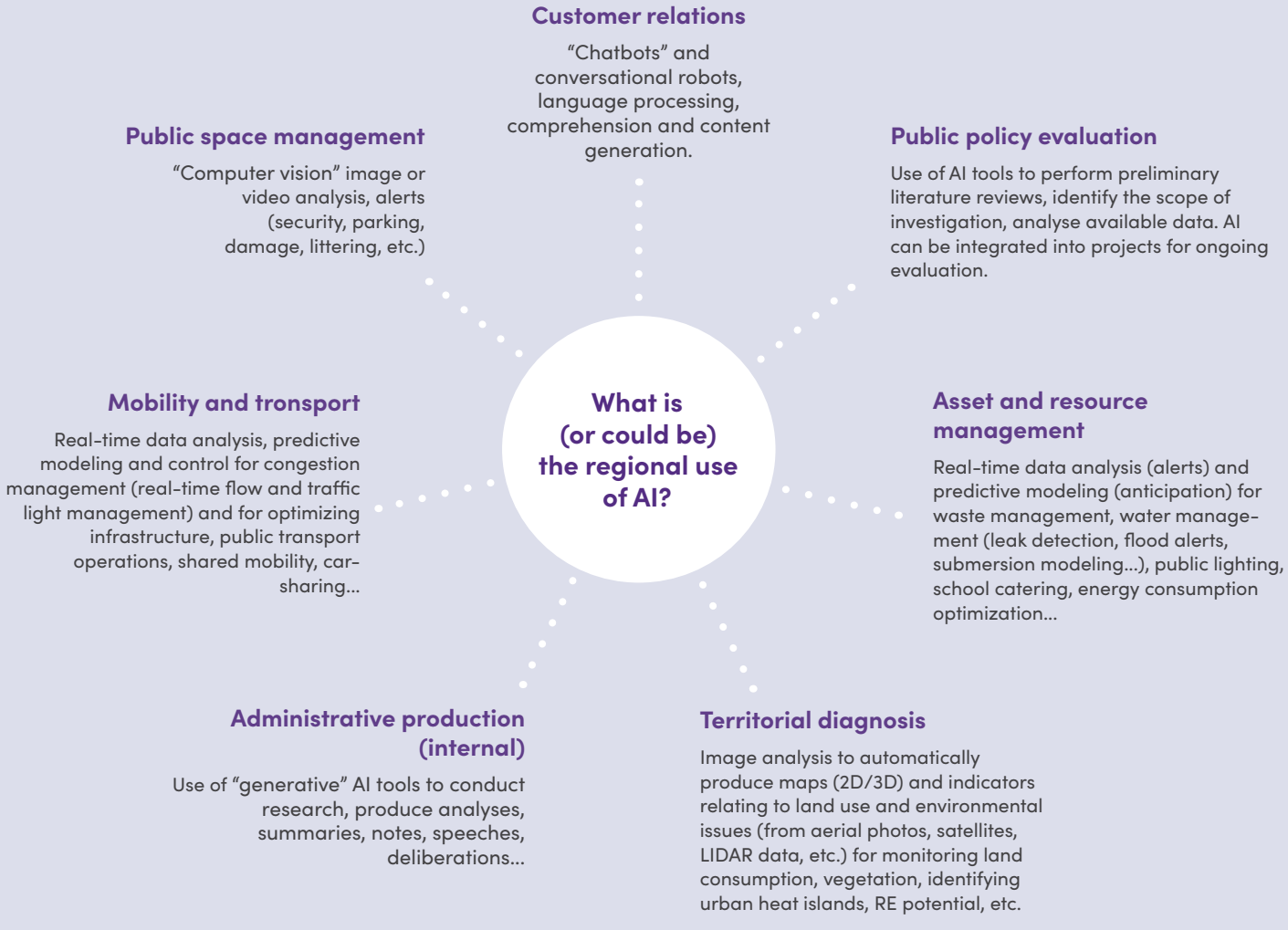
5 French Council of State (2022). *Intelligence artificielle et action publique : construire la confiance, servir la performance*. Available online: < https://www.conseil-etat.fr/publications-colloques/etudes/intelligence-artificielle-et-action-publique-construire-la-confiance-servir-la-performance#:~:text=Intelligence%20artificielle%20et%20action%20publique%20%3A%20construire%20la%20confiance%2C%20servir%20la%20performance,-31%20août%202022&text=Si%20l%27intelligence%20artificielle%20suscite,impossibles%20à%20réaliser%20jusque%2Dlà.>

6 European Commission (2021). *Artificial intelligence in public services*.

But the type of project which is not only the most common, but also the fastest and seen as the easiest to deploy, is based on generative AI. These projects being launched at the initiative of public services. But they are often in response to solicitations from large digital companies or local start-ups. Regional LLM projects are multiplying. DelibIA, for example, is a project for generating local legal texts, based on the intake of millions of existing municipal deliberations. Numerous conversational robot projects are being developed to guide users through the maze of public assistance. The French government is itself experimenting with these tools and is developing its own language model. This open-source "sovereign administrative ChatGPT" is called ALBERT. It is available to all government agencies.

Finally, CIVITEO and the Data Publica Observatory have identified seven major areas of possible applications of AI within local public administrations.

Diagram 1: Possible areas for the use of AI in local public service missions.



But the use of these tools is not straightforward. There are many economic, legal/ethical, environmental, technological and even democratic and political issues at stake. These tools are available in the corporate world, but their impact is significantly different when used for the general interest. This raises important questions for public decision makers.

II. Useful AI: building a decision-making tool for local public decision makers

The tool presented here is the fruit of concrete work carried out by CIVITEO teams with local decision makers from various local authorities. We have thus been able to work on the development of several projects involving the use of AI on a regional scale, including the PrevizO project run by the Climate Data Hub of the Centre Val-de-Loire Region (a tool for modelling water scarcity in a Loire watershed), and the Urba (IA) project of the Paris-Saclay agglomeration (a combination of algorithms for modelling the impacts of urban planning rules on climate, biodiversity, soil artificialization, etc.).

We were also able to test and co-construct this decision-making tool by mobilizing the municipal teams of several communes. The city of Angers, the first French metropolis to vote for a "metropolitan data strategy", wanted to develop an internal policy on the use of AI in its public service missions. The towns of Montreuil and Clichy-sous-Bois have organized AI seminars for their executives on this topic. The city of Montpellier has set up a test panel of volunteer employees representing a wide variety of professions as part of a more global initiative to discuss artificial intelligence in the service of the general interest.

A) The need to support reflection

Each artificial intelligence system likely to be used by the local public service is characterized by a multitude of criteria likely to be taken into account in the decision as to whether or not to use it.

Is the algorithm open or not? Has it been developed by a French or European company? Does it meet (by anticipation) the criteria in future European AI Act regulations? What do we know about the training data? If it's a specific AI, has it been trained with data from French cities? Or American, or Chinese? Does the use of the algorithm guarantee compliance with the GDPR or, on the contrary, weaken the conformity of the service being rendered? How is the algorithm's performance monitored? How are biases detected? Who has the capacity to correct them, to perfect the training? What is the environmental impact of this AI system? Can we measure it? What impact will it have on civil service jobs?

There is an abundance of literature, charters, and guides on all these points. In France, the Conseil d'Etat has published a report[7] in which it lists the conditions for a trusted AI at the service of public action: human primacy, strategic autonomy, environmental sustainability, security, transparency, equity and non-discrimination, and performance.

But this theoretical framework is complex to apply for local administrations who are faced with AI opportunities whose promises are sometimes very tempting, and who might not have the necessary to respond or even identify the issues when faced with real AI risks.

This is why the analysis grid developed with a few pioneering local authorities aims to balance elements that can be grasped locally by teams trained in at least the key concepts of artificial intelligence and the legal framework for their deployment (in France, for example, there is a little-known and little-enforced law on algorithmic transparency, which is particularly strict when AI is used). The aim is not to be "for" or "against" the use of AI for a socially useful project, but to be able to say that it can be deployed "provided that", or that it shouldn't be deployed "because".

And for every project, the first question is the usefulness of AI.

7 French Council of State (2022), *supra.*

B) A tool for local authorities

The examination of an AI project begins with an estimate of the expected benefits. These are of three kinds. Firstly, functional: AI can improve forecasts, reduce processing times, eliminate repetitive tasks, and generate substantial savings. The benefits can also be linked to a region's strategy, its attractiveness, and its eco-system. They can, of course, be political, in the sense that the use of AI supports a priority policy supported by elected representatives.

There are also risks, and therefore, conditions to be met to deploy the system with confidence. In the grid drawn up with local authorities, we have identified three risks. The first risk concerns the control and monitoring of biases and errors (human supervision being one of the possible conditions). The second concerns the control of the algorithm and data (enabling strategic autonomy and a state of transparency), and the third concerns the protection of privacy (with the possibility of in-depth impact studies to guarantee compliance with the GDPR).

In addition to these six criteria of opportunities and risks, there are 2 criteria whose impact can be measured on a balance scale: environmental impact and job impact. As far as environmental impact is concerned, the emergence of high-performance tools that measure the carbon impact of algorithms will make decision makers' work easier. The notion of frugal AI is strongly promoted in France by the Ministry of Ecological Transition[8] and is based on a simple idea: an AI system that creates more problems than it solves must be abandoned. Overall assessment is complex, and in case of doubt, if no frugal model exists, the recommendation may be to abandon AI.

When it comes to the impact on jobs, the process is equally complex. Some pioneering local authorities in France are using an adapted version of the "Measuring the Social Acceptability of AI in the Workplace" project developed by ICAM and Yann Fergusson as part of the Global Partnership on Artificial Intelligence.[9] The aim is to move away from simplistic arguments concerning AI as a factor in the immediate and massive loss of jobs, and to measure its impact on several factors of professional and social acceptability. In particular, for local public services, the possible impact of AI on the relationship between users and public agents.

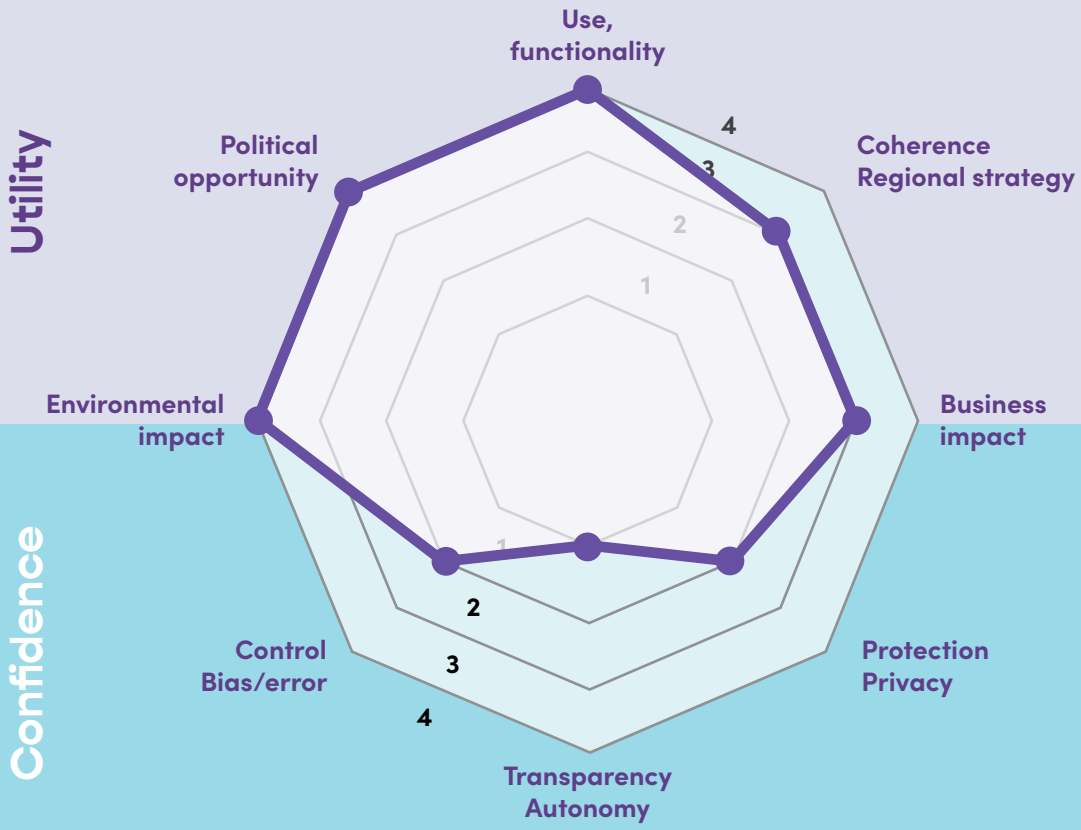
For the environmental and job criteria, the goal is to achieve a positive impact while minimizing negative effects. Hence the concept of a balance.

In very concrete terms, the tool devised by CIVITEO with pioneering local authorities is based on a four-level analysis grid: low utility/confidence; moderate utility/confidence; medium utility/confidence; and high utility/confidence. It is represented in the form of a radar chart, which ultimately enables local authorities to assess the impact that an AI system could have on the implementation of a given public service mission.

8 Ministry of Ecological Transition and the Cohesion of French Territories (2023). *Appel à projets Démonstrateurs d'IA frugale pour les transitions écologiques et énergétiques dans les territoires.*

9 CRESPIN, P. and *Al.* (2022). *L'IA au travail : propositions pour outiller la confiance,* Conférence Nationale sur les Applications Pratiques de l'IA. June 2022, Saint-Etienne, France.

Diagram 2: The radar chart for regional AI systems.



On this radar chart, the proposed solution may seem very attractive in terms of immediate benefits, political display, and strategy, but the AI system offers very few guarantees in terms of transparency, public control, compliance with the GDPR in terms of discrimination risks, etc. It may well be a very fashionable algorithm, and the temptation to succumb to the hype is strong!

Of course, no AI system is optimal. There are always checks and guarantees that need to be provided. Once the study has been carried out, two scenarios are possible: "the local authority decides to experiment with the use of AI in a specific service, and here is how" or "the local authority decides to abandon this AI project, for the time being, and here's why...".

Vigilance is essential. Some AI systems fail to provide the necessary assurances for their reliable deployment for projects of general interest. Others, however, do offer these guarantees ... and turn out to be unnecessary.

European Health Data Space: Issues and Challenges for the Secondary Use of Health Data

Nathalie De Grove-Valdeyron

Professor of Public Law, Toulouse Capitole University (IRDEIC)
Jean Monnet Chair (2023-2026)

Ursula Von der Leyen has made the digital world one of the main priorities of her mandate, believing that "data and AI[1] are the ingredients for innovation that can help find solutions to today's societal challenges [in particular] in healthcare [...]". Back in 2018, the Commission identified major obstacles to the adoption of digital solutions in healthcare in its communication on the digital transformation of care. Priorities already included, on the one hand, "secure access for citizens to their health data and its sharing" and, on the other, "the quality of this data to promote research, disease prevention, and personalized care". These two priorities are the two main objectives of the European Health Data Space (EHDS), which was the subject of a proposal for a regulation presented by the European Commission on May 3, 2022 (hereafter referred to as the EHDS Regulation). [2] Furthermore, in its communication on February 19, 2020 (COM [2020]66), the Commission proposed a strategy for data that took shape in various "Acts", which are in reality regulations. They need to be considered to understand both the context in which the new regulation will fit and the still uncertain relationship between the EHDS Regulation and these various texts. While we cannot delve deeply into this question within the scope of this paper, we will briefly highlight a few key texts that are inseparable from the proposed regulation. First of all, there is the "Data Act", the data regulation of December 13, 2023,[3] which implements certain objectives of the "Data Governance Act", Regulation (EU)2022/868 on data governance.[4] Then, there is the "Artificial Intelligence Act" of April 21, 2021, which was the subject of a political agreement on December 8, 2023, adopted unanimously by the governments of the member states in January 2024. These regulations have a cross-cutting dimension and are intended to apply to the health sector in certain aspects.

1 See Commission Communication of February 20, 2020, COM (2020)65 final, White Paper: *Intelligence artificielle, Une approche européenne axée sur l'excellence et la confiance*. This White Paper accompanies the European Data Strategy.

2 COM (2022),197 final of May 3rd, 2022.

3 Regulation (EU)2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonized rules on fair access to and use of data, OJEU L 22 December 2023.

4 Example of articulation: According to recital 115 [...] This regulation (Data Act) is "without prejudice to more specific rules in the context of the development of common European data spaces or, subject to the exceptions provided for in this Regulation, to Union and national law providing for access to and authorising the use of data for scientific research purposes". See, for example, the request to make personal data available under article 14 for exceptional need (like emergency) and the processing of data in this context (art.17, 1 point g and h) (pseudonymization/anonymization).

This is particularly true for rules related to the voluntary sharing of data, such as those established by data altruistic organizations recognized under Regulation (EU) 2022/868 of May 30, 2022 (strictly regulated),[5] or for rules on the provision by certain data holders of personal data in cases of exceptional need. [6] Regarding AI, the European Union has emphasized the importance of AI systems applied to large datasets, particularly for research and the development of increasingly precise and personalized medicine. Among the other regulations that will need to be coordinated with the EHDS Regulation, it is also necessary to specifically mention the Medical Devices Regulation (EU)2017/745 and In Vitro Diagnostic Devices Regulation (EU)2017/746 of 2017 and, it goes without saying, the Data Protection Regulation (EU) 2016/679 of April 27, 2016 (GDPR)[7] as well as Regulation (EU) 2018/1725 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies.[8] A clarification of the interactions between these various secondary laws would certainly be welcome to avoid divergent interpretations during the concrete implementation of the regulation by the Member States. Directive 2011/24/EU on the application of patients' rights in cross-border healthcare is also noteworthy in that it was behind the creation of the eHealth network, which brings together, on a voluntary basis, representatives from member states (national experts in online healthcare). This network has been the primary strategic and governance body for electronic health data interoperability issues. However, its role, limited to a supporting role and the adoption of non-binding acts, has not enabled it to achieve its objectives. This was one of the factors that led to the proposal of the EHDS Regulation.

It is certainly too early to assess the concrete impact that the creation of this new common data space in the health sector will have, particularly for the secondary use of health data, at the current stage of the regulation's development, as these lines are being written.[9] Nevertheless, by deliberately stepping away from traditional French academic norms, we find it interesting to examine the stakes of this new space. What will the EHDS bring, particularly in terms of the secondary use of health data (II), and how will this data be made available, given the respective competences of the European Union and the Member States? To answer these two questions, we first need to recall why the creation of such a data space was necessary (I).

I. Why create a European health data space?

The aim of the EHDS regulation is ambitious: The aim is to ensure that "individuals in the Union have greater control over their health data in practice," but also to "establish a legal framework consisting of reliable governance mechanisms at both the Union and Member State levels and a secure processing environment". The construction of this common data space should also contribute to the establishment of a single market for digital health products and services by facilitating the cross-border flow of electronic health data. Article 114 of the TFEU has therefore naturally been chosen as the legal basis for the regulation, since its aim is to improve the free movement of goods and services while taking account of a high level of health protection in order to take advantage of the internal market, and thus put an end to obstacles to access and use of electronic health data. This legal basis is accompanied by Article 16 of the TFEU (right to protection of personal data such as health data).

5 Hulin A-S. and Castets Renard C. (2022). *Le règlement sur la gouvernance des données ou la promesse d'un marché des données à l'européenne*, Dalloz IP/IT, 575.

6 See, for example, the request to make personal data available under article 14 for exceptional needs (emergency situation) and the processing of data in this context (art.17, 1 point g and h) (pseudonymization/anonymization).

7 OJEU L 119, May 4, 2016, p. 1.

8 OJEU L 295, November 21, 2018, p. 39.

9 At the time of going to press, the proposal was adopted by the European Parliament at first reading on 24 April 2024. See <https://www.europarl.europa.eu/doceo/document/TA-9-2024-0331_EN.html> (document P9-TA (2024)0331). The text is awaiting the Council's position at first and only reading.

The proposed regulation distinguishes between two types of healthcare data use: "primary" and "secondary". The "primary" use is intended to ensure continuity of care and improve patient management in a cross-border context. The studies commissioned by the Commission show that this continuity is far from being achieved, even at Member State level. In particular, the diversity of standards and technical specifications used (incompatible software between primary care and hospitals, unusable operating systems, etc.) has been highlighted. As data interoperability and portability, which are a prerequisite, are not satisfactorily ensured at national, or even regional and local levels, it is easy to imagine the amplification of problems in a cross-border context. This observation underlines the ambitious nature of the Commission's proposal, if only in terms of the technical challenge it represents.

The European health crisis related to COVID-19 highlighted the need to strengthen European coordination to establish a concerted policy in response to health emergencies. This coordination enabled the implementation of technical solutions ensuring data transferability and interoperability while respecting confidentiality. According to the experts, the challenges posed by the technical aspects can now be met. The Union's digital Covid certificate is undoubtedly the most tangible example[10] of the fruits of this European cooperation and the benefits it has brought in terms of facilitating freedom of movement within the Union. Today, the secure digital platform MyHealth@EU,[11] initially known as "e-HDSI", has been designed to enable the cross-border transfer of health data as part of the primary use of health data. It now awaits use by all Member States to "streamline" the healthcare journey within the Union, at the pace set by the Member States, as the use of this platform is currently voluntary. The platform already connects Member States via their national contact points (NCPe-H) (National Contact Point for eHealth),[12] but it is awaiting the adoption of the relevant regulation before being able to deploy its full potential. By 2025, it should enable the transfer of electronic prescriptions, electronic medical records, hospital discharge reports, laboratory reports, and medical imaging reports, using a European health data exchange format. The so-called secondary use of health data appears to be an even more sensitive issue.

II. What does the regulation do about the secondary use of health data?

Secondary use of data, the second objective of EHDS, is presented in the Commission's proposal as (use) "[...] for other purposes that would benefit the society such as research, innovation, policymaking, patient safety, personalized medicine, official statistics or regulatory activities'". It finds its legal basis for processing in Article 9§ 2 points g, h, i, and j of the GDPR.

The idea is to enable researchers, innovators, policymakers, and regulators to access quality data for their work securely, with reliable governance, and at a lower cost than a consent-based system.

10 Regulation (EU) No. 2021/953 of June 14, 2021, on a framework for the issuance, verification and acceptance of interoperable COVID-19 vaccination, test and recovery certificates (EU Digital Covid Certificate) to facilitate free movement during the Covid-19 pandemic, *OJEU* No. L 211, June 15, 2021, pp. 1-22.

11 This platform was called E-HDSI (eHealth Digital Service Infrastructure) see, De Grove-Valdeyron N. (2022). « Santé Publique », *Annuaire de droit de l'Union européenne*, dir. C. Blumann and F. Picod, Panthéon Assas, 847 à 876, online: < https://www.cairn.info/annuaire-de-droit-de-l-union-europeenne--9782376510567-page-847.htm>.

12 In France, the *Agence du numérique en santé*.

The very notion of secondary use has given rise to debate and recommendations by the TEHDaS joint action.[13] It is defined in the regulation as use for scientific research for purposes strictly mentioned in the regulation (see below), but it implies the possibility of re-using data for purposes other than those for which they were initially collected, which raises the important question of consent, or at the very least, of informing patients about the use of their data for research purposes. In France, the question of consent was the subject of an opinion issued on May 9, 2023,[14] by the CCNE (*Comité Consultatif National d'Éthique*) and the CNPEN (*Comité National Pilote d'Éthique du Numérique*). This opinion, which includes numerous recommendations, sets out the different types of consent that could be required in the context of health data platforms: general consent (absence of knowledge of the specificities of studies at the time of the request), dynamic consent (via a specialized online communication platform that facilitates consent between researchers and participants), specific dynamic consent (possibility of specifying authorized accesses), and meta-consent (consent options according to project models and forms), which is a model used in Canada by the GRIIS platform.[15] Today, the way in which consent is collected at members State level is not subject to common provisions. For example, in the case of the creation of the *Dossier médical partagé* (shared medical file), which got off to a rocky start in France, the "Opt out" technique was adopted,[16] enabling the widespread use of "*mon espace santé*" (my health space), with 98% of insured persons benefiting from this online space by February 2023 (unless they objected to its creation).

It should also be noted that the way in which consent is obtained is not currently covered by European provisions, as this is a matter concerning the organization of healthcare systems and therefore falls within the competence of the Member States in accordance with Article 168§ 7 of the TFEU. This issue of consent is one of the particularly sensitive points of the proposal. The European legislator seems to be moving towards allowing Member States some leeway in this area, considering their national particularities. However, it mandates the implementation of mechanisms that enable individuals to exercise their rights and to reconsider a refusal of consent.[17] With the support of the e-health network, the Commission is also considering the introduction of a "harmonization"[18] in the processing of consent, along the lines of a single form enabling consent to be processed in a similar way in all member states. Article 22 of the "Data Governance Act" also calls on the Commission to adopt delegated acts forming a "set of rules" laying down, in the specific context of data altruism, "appropriate information requirements to ensure that data subjects and data holders are provided, before consent or permission for data altruism is given, with sufficiently detailed, clear and transparent information regarding the use of data, [and] the tools for giving and withdrawing consent [...]" (art.22§ 1, a and for the relevant purpose, see art.21§ 1 a)). This could involve health data, and these delegated acts could, in our view, serve as inspiration for the collection of consent in the context of secondary use of health data, if some form of consent is required by a Member State in accordance with its national law.

13 TEHDaS (Towards European Health Data Space), (2023). *Qualitative Study to Assess Citizens' Perception of Sharing Health Data for Secondary Use and Recommendations on How to Engage Citizens in the EHDS*, deliverable 8.1, online: <https://tehdas.eu/tehdas1/app/uploads/2023/03/tehdas-study-to-assess-citizens-perception-of-sharing-health-data-for-secondary-use.pdf>.

14 CCNE and CNPEN, (2023). *Plateformes de données de santé : enjeux d'éthique*. Avis 143 du CCNE, Avis 5 du CNPEN, online: <https://www.ccne-ethique.fr/sites/default/files/2023-05/CCNE-CNPEN_GT-PDS_avis_final27032023.pdf>.

15 See Barton A., Cumyn A. and Ethier J.-F, *Le métaconsentement en lumière*, online:< https://ssaquebec.ca/wp-content/uploads/2018/05/Cumyn_A.pdf>. See also Barton A., Cumyn A. and Ethier J.-F., *Atelier B3 : Le métaconsentement : point de citoyen sur une approche alternative au consentement*, online:<https://www.msss.gouv.qc.ca/professionnels/documents/comites-d-ethique-de-la-recherche/journees-detude/2019-9e-journees/13h30-24oct_B3-Cumyn_JECER2019.pdf>.

16 See <https://www.ameli.fr/assure/protection-donnees-personnelles>.

17 According to the latest available version, *" natural persons shall have the right to opt-out at any time and without stating reasons from the processing of personal electronic health data relating to them for secondary use* [...]". Member States shall provide for an accessible and easily understandable opposition mechanism for exercising this right (Article 48a (1)), but Member States retain the possibility of making certain data available, despite an opposition (Article 34(1)a) to c) (Article 48a(3)).

18 The term "harmonization" is not to be understood here in the legal sense of Article 114 of the TFEU, but rather in the sense of "approximation".

III. Secondary use of health data: what data for what purpose?

The implementation of the EHDS implies "governance", and therefore the adoption of rules enabling data to be made available under the conditions laid down by the regulation. Without being able to analyze this governance here, we will focus on two delicate points in this context: the data concerned and the purpose of processing it. Firstly, only certain categories of electronic (health) data are targeted for secondary use (see art.33). These include, but are not limited to, data from electronic medical records (EMRs); administrative data relating to claims and reimbursements; genetic, genomic, and proteomic data; data generated automatically by medical devices, etc. As for the only authorized purposes, they are listed in article 34,[19] these include a) *reasons of public interest* in the field of public health and occupational health (such as activities designed to protect against serious cross-border threats to health and to monitor public health, or activities designed to guarantee a high level of quality and safety in healthcare, including the safety of patients and of medicinal products or medical devices); b) policy development and regulatory activities to support public sector bodies or Union institutions, agencies and bodies including regulatory authorities, in the health or care sector to carry out their tasks defined in their mandates; this also includes c) producing official statistics relating to the health and care sector; or e) conducting scientific research relating to the health or care sectors, contributing to public health or health technology assessment (or guaranteeing a high level of quality and safety of healthcare, of medicinal products or of medical devices), including the development and innovation of products or services; training, testing, and evaluation of algorithms, AI systems, and digital health applications...

The regulation then specifies the obligations of the organizations responsible for data access, designated by the Member States, which must provide users with "relevant" data in a secure processing environment. These organizations maintain a national catalog of data sets (including source, nature, and conditions of availability) and are subject to specific obligations detailed in the regulation.

Data holders must, upon request, make the requested data available to applicants for the previously indicated purposes within a reasonable period limited to 3 months from the receipt of the request. However, this timeframe is currently considered impractical, as the data availability period in France is closer to 18 months.

The applicant must provide a detailed explanation of the intended use of the health data, as well as protect against the risk of re-identification, etc. A single application can be submitted when it concerns more than one Member States (via the dedicated HealthData@EU secure platform). Member States and the Commission are working to ensure the interoperability of health data with other future common data spaces.

While the main rules, briefly mentioned above, have been established in the regulation, it is clear that its concrete implementation and effectiveness will depend on numerous delegated and implementing acts entrusted to the European Commission. Until these are adopted, the effective application of the regulation risks being delayed by several months or even years, as several uncertainties remain at this stage. For the time being, a new TEDHaS 2 action has been set up, which should help Member States in anticipating any problems that may arise upstream.

19 Considering the amendments proposed by the European Parliament.

Only a robust framework for the governance of health data will maintain public trust in this new European Health Data Space, which should enable the full potential of digital health to be realized. Trust in this EHDS also requires informing patients/citizens and gaining their consent to share essential health data for research purposes (within the scope of the so-called "secondary" use of health data). The right balance needs to be struck between the public interest in exploiting vast databases of health data[20] and the dizzying prospects that may arise from such exploitation at the level of the European Union (notably by applying artificial intelligence systems), on the one hand, and the safeguarding of private interests, on the other. This delicate balance should incorporate respect for the principle of proportionality[21] into the balance, bearing in mind the possible recourse by the Union or by Member States to Article 89§ 2 of the GDPR which provides the possibility of derogations in the case of scientific research under certain conditions.

The participation of citizens and all stakeholders, particularly through the EHDS Committee and the stakeholder forum, should strengthen trust in this European health data space. Existing health data infrastructures, set up within a national regulatory framework, are already encouraging research and innovation activities in the health sector, but they now need to be designed in a cross-border context with appropriate governance mechanisms to make the best possible use of the databases for the purposes set out in the regulation. Only in this way can major advances be made in European health research.

20 Supiot E. (2022/1). « Les données génétiques au service de la recherche scientifique », Droit, santé et société, n° 1, 53-58.

21 It is also a question of avoiding excessive costs or cumbersome formalities linked to the collection of consent, which could be introduced by certain member states, thereby undermining the coherence of the regulation and reintroducing a new legal fragmentation of data access, which is precisely what the regulation aims to combat.

Data and AI in healthcare: what are the new challenges?

Anita Burgun

MD, PhD, Professor of Medical Informatics at Université de Paris. Chair of the Department of Medical Informatics at Georges Pompidou and Necker Hospitals (AP-HP).

Introduction

With the exception of medical imaging and laboratory results, most of the information we refer to as health data is entered by healthcare professionals. They often experience this time spent on entering patient data as a workload that reduces their availability for more medical tasks. For emergency physicians at London's Whittington Hospital, for example, the time taken to enter data for a patient seen in the emergency department increased more than 3-fold when they switched from paper to computerized records[1]. A recent study carried out in the USA in ambulatory care and adult medicine, covering 100 million consultations by 155,000 doctors, showed that a doctor spends an average of 16 minutes and 14 seconds per consultation on his or her computer, interacting with the clinical information system. Most of this time is devoted to analyzing the patient's file (33%), documenting the file (24%) and computerized prescribing (17%). In total, she/he spends between 3.5 and 6 hours per day this way, with a median of 4:01 a.m./2. [2].

The arrival of Artificial Intelligence (AI) in healthcare has further increased the "pressure of data," leading to the belief that all we need to do is make data available to create beautiful and useful models to manage patients, and some proponents of "technological solutionism" have naively imagined that we could replace clinicians with artificial intelligence models for certain tasks, ignoring the issues of data production and medical decision-making. In contrast to these simplistic positions, American medical scientific societies are currently warning of the need to redefine the priorities of AI in healthcare. In November 2023, the American Medical Informatics Association (AMIA) joined forces with the Association of Medical Directors of Information Systems and the Alliance for Nursing Informatics, to assess AI's ability to reduce clinicians' workloads and limit the burnout generated by computerization. The aim of AI in this context is to lighten the burden of research and data production in patient records. Also in 2023, the American Medical Association (AMA) called for clinicians to have a leading voice in defining what role AI should play in medicine. The AMA also insisted on the need for mature clinical information systems as a prerequisite to the implementation of any AI and to evaluate the benefits to clinicians of real-life AI.

1 Armstrong S. (2017). "Data, data everywhere: the challenges of personalised medicine", *BMJ*, 359, online: <https://www.bmj.com/content/359/bmj.j4546>.

2 Overhage J-M. and McCallie D. Jr (2020). "Physician Time Spent Using the Electronic Health Record During Outpatient Encounters: A Descriptive Study", *Ann. Intern. Med.*, vol. 172-3, 169174, online: <https://pubmed.ncbi.nlm.nih.gov/31931523/>.

Our aim is to build the framework for implementing Learning Health Care Systems, which move away from the schematic vision of healthcare data reuse for research by proposing an intrinsically translational objective. Specifically, we focus on three points: taking into account the multimodal characteristics of patient data, federated approaches to sharing, and clinical decision support in learning systems.

I. Multimodal Approaches

rather than re-entering data as done in clinical trials. The patient record is a source of structured data in the form of codes from a thesaurus, data produced by monitors (pulse, blood pressure, etc.) or laboratory results, and unstructured data, whether images or text documents. The individual data present in clinical information systems describe the patient's condition at the time of examination, his or her clinical history, and family history. It is multimodal in nature, including various medical imaging modalities, text reports, results from biological and "omic" analyses, continuous measurements, therapeutic prescriptions (drugs, procedures) and other structured data.

The rich, multimodal nature of this raw data makes it ideal for training Artificial Intelligence models[3] and opening up new opportunities for applications[4]. Although most AI models so far have been trained to use a single data source, typically imaging (with the term multimodality referring to the combination of sources like CT and MRI scans), and for a specific task such as classifying lesions as malignant or benign, the current challenge is to move beyond single-task and single-modality models to incorporate as many data sources as possible. For example, it is useful to integrate genomic data, medical imaging, and other clinical data from the patient file to make therapeutic choices in precision oncology; biological, clinical, imaging, and connected object data to create digital twins.

Rare genetic diseases are a medical field in which it is particularly important to integrate as much data as possible to establish associations between phenotypes and genotypes, and to reuse the knowledge extracted from these data in the care of new patients. Research at the Imagine Institute focuses on genetic diseases, and the Institute works very closely with the *Hôpital Necker Enfants Malades* of the *Assistance Publique-Hôpitaux de Paris*. There are some 8,000 rare genetic diseases in all, and each is characterized by a very low incidence. Each case must be described with the utmost precision to achieve the most comprehensive understanding possible. All sources of information must therefore be integrated, including all clinical data in the patient file and all available genetic data.

3 Acosta J. N. and Al (2022). "Multimodal biomedical AI", *Nat Med*, vol. 28-9, 17731784.

4 Topol E. J (2023). "As artificial intelligence goes multimodal, medical applications multiply", *Science*, vol. 381, 6663, online: <https://pubmed.ncbi.nlm.nih.gov/37708283/>.

II. A Federated approach

An accurate representation of the patient cannot be considered free from the risk of re-identification. An example is syndromic intellectual deficits with facial dysmorphia, where facial photographs are necessary to establish the diagnosis. The data storage infrastructure and the data governance model must therefore be adapted so as to optimally meet the combination of the following two desiderata:

1. The data of patients with rare genetic diseases is extremely sensitive and must be ultra-protected;

2. For a given rare disease, the number of cases per centre, or even per country, can be extremely low, and only by pooling resources can significant advances in research be made.

An approach based on distributed databases and federated learning, in which the data remains local but the models are shared between institutions, is the solution that makes it possible to respond to these contradictory injunctions[5]. The aim is for all participating centres to harmonize their data by adopting a common data model[6] so that the same queries can be used by all centres and so that the algorithm can be trained in each centre and then shared.

In a federated system, data remains hosted and controlled locally, and is not exported either to other sites or to a central data hub. The federation of (distributed) databases provides better guarantees against possible breaches of confidentiality and facilitates control by data producers, who can decide whether or not to participate in a project. As a result, collaboration between centres is strengthened, internationalization is facilitated, and public confidence is reinforced or preserved.

The federated approach means that multimodal data produced at site level is available without having to store it in a mega-database collecting individual data from several sites and several modalities. This makes it easier to recruit new data sources whether modalities or sites.

Finally, federated data analysis relies on better control of the data production pipeline and the resulting quality. This control of the data production and analysis chain is likely to produce better AI models. Local validation is necessary to verify the transposability of external models,[7] and is facilitated by this distributed approach, in which each site is responsible for validation.

The federated approach has been adopted by international networks, for example during the COVID crisis by the Consortium for Clinical Characterization of COVID-19 by EHR (4CE). Right from the start of the crisis, this approach enabled international research into COVID to be carried out based on the reuse of data routinely produced by hospitals, in line with the WHO's recommendation to exploit healthcare data[8]. This approach has enabled us not only to generate new knowledge about the disease, but also to generate predictive models and, above all, to assess their transposability/generalizability[9].

More generally, the distributed/federated approach is advocated at national level by countries such as Germany and Canada. The German Miracum network[10] brings together academic, hospital and industrial partners to create local data centres, using principles of interoperability and data harmonization (in particular, the creation of a common data model) that make it easy to share database query tools and AI models[11].

---

5 Sheller M. J. and *Al.* (2020). *Scientific Reports,* vol. 10, 12,598, online: <https://www.nature.com/articles/s41598-020-69250-1>.

6 See Kaliyaperumal R. and *Al.* (2022). "Semantic modelling of common data elements for rare disease registries, and a prototype workflow for their deployment over registry data", *Journal of Biomedical Semantics,* vol 13, art. 9, online: <https://jbiomedsem.biomedcentral.com/articles/10.1186/s13326-022-00264-6>.

7 Youssef A. and *Al.* (2023). "External validation of AI models in health should be replaced with recurring local validation", *Nature Medecine,* vol. 29-11, 26862687, online: <https://www.nature.com/articles/s41591-023-02540-z>.

8 Brat G. A. and *Al.* (2020). "International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium", *NPJ Digit. Med.*, vol 3–109, online: <https://pubmed.ncbi.nlm.nih.gov/32864472/>.

9 Weber G.M. and *Al.* (2022). "International comparisons of laboratory values from the 4CE collaborative to predict COVID-19 mortality", *NPJ Digit. Med.*, vol 5–74, 18, online: <https://www.nature.com/articles/s41746-022-00601-0>.

10 Prokosch H-U. and *Al.* (2018). "MIRACUM: Medical Informatics in Research and Care in University Medicine", *Methods Inf. Med.*, vol. 57-S 01, online: <https://pubmed.ncbi.nlm.nih.gov/30016814/>.

11 Kapsner L. A and *Al.* (2021). « Linking a Consortium-Wide Data Quality Assessment Tool with the MIRACUM Metadata Repository », *Appl. Clin. Inform.*, vol. 12-4, 826835, online:< https://pubmed.ncbi.nlm.nih.gov/34433217/>.

Towards a Learning Healthcare System

Researchers, doctors, regulators, and public authorities have become accustomed to separating the healthcare system from research: the healthcare system produces health data as part of patient care and this data can then be used for research. In this dichotomous view, the challenge lies in how to provide researchers access to the data while ensuring security, confidentiality, and adherence to good practices. This has given rise to ministerial missions aimed at laying the foundations of a roadmap for the re-use of healthcare data and has led to reports such as the recent *Fédérer les acteurs de l'*écosystème *pour libérer l'utilisation secondaire des données de santé* in France. This report recommends streamlining regulations and accelerating the availability of data for research[12].

However, it is desirable to propose an alternative and more integrated vision based on the model of the learning healthcare system, so that the "final leg of the race," which aims to enable clinicians to benefit from the products generated by making healthcare data available, is considered.

Let's take the example of the *Hôpital Necker Enfants Malades*, part of *the Assistance Publique Hôpitaux de Paris (AP-HP)*. The Necker Hospital is associated with a research institute for genetic diseases, the Imagine Institute. The latter is developing research into genetic diseases, and its researchers are producing "mechanistic" biological models of these diseases based on "wet lab" experiments. Patient phenotyping is carried out using the clinical data warehouse at the Necker Hospital. Similarity measures on patient phenotypic profiles are used to identify patient subtypes in pleiotropic diseases[13]. The same methods are useful in the clinical setting for an undiagnosed patient to find similar patients in the data repository for whom a diagnosis has been established.[14]. In this case, the reuse of data from the health warehouse is used to reduce the time spent determining a diagnosis and to better manage new patients admitted to hospital[15]. Figure 1 illustrates the concept of a learning healthcare system as developed at the Imagine Institute and the Necker Hospital. Its value to clinicians has been demonstrated (one third of the warehouse's one million patient records in 2024 have been "touched" by a query), as has its ability to speed up diagnosis[16].

12 Marchand-Arvier J. and *Al.* (2023). *Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire des données de santé,* Inspection Générale des Affaires Sociales, 2023, online : <https://sante.gouv.fr/IMG/pdf/rapport_donnees_de_sante.pdf>.

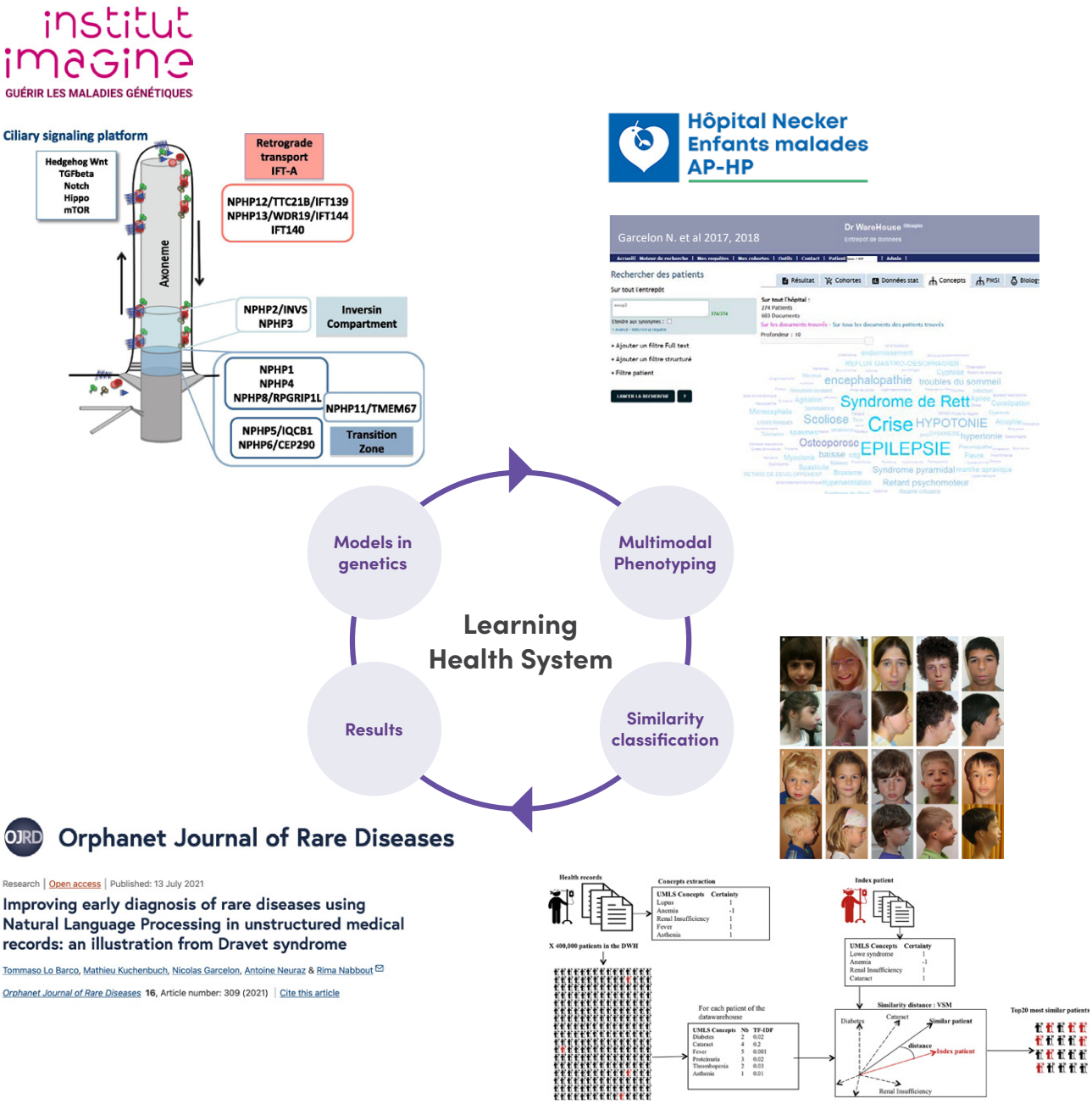
13 Chen X. and *Al.* (2019). « Phenotypic similarity for rare disease: Ciliopathy diagnoses and subtyping », *J. Biomed. Inform.*, vol. 100, online: <https://pubmed.ncbi.nlm.nih.gov/31622800/>.

14 Lo Barco T. and *Al.* (2024). "Natural history of rare diseases using natural language processing of narrative unstructured electronic health records: The example of Dravet syndrome", *Epilepsia,* vol. 65-2, 350361, online: <https://pubmed.ncbi.nlm.nih.gov/38065926/>.

15 Faviez C. and *Al.* (2024). « Performance and clinical utility of a new supervised machine-learning pipeline in detecting rare ciliopathy patients based on deep phenotyping from electronic health records and semantic similarity », *Orphanet. J. Rare Dis.*, vol. 19-1, 55, online: <Performance and clinical utility of a new supervised machine-learning pipeline in detecting rare ciliopathy patients based on deep phenotyping from electronic health records and semantic similarity>.

16 Lo Barco T. and *Al.* (2021). « Improving early diagnosis of rare diseases using Natural Language Processing in unstructured medical records: an illustration from Dravet syndrome », *Orphanet. J. Rare Dis.*, vol. 16-1. 309, online: <https://pubmed.ncbi.nlm.nih.gov/34256808/>.

Figure 1: Imagine Institute and Necker Hosptial's learning system for genetic diseases.



Sources :

https://www.sciencedirect.com/science/article/pii/S153204641930228X#f0020.

https://ojrd.biomedcentral.com/articles/10.1186/s13023-021-01936-9.

https://www.has-sante.fr/upload/docs/application/pdf/2019-11/colloque-has-a-burgun.pdf.

https://www.researchgate.net/figure/Facial-features-of-individuals-with-Meier-Gorlin-syndrome-Frontal-and-lateral-view-of-10_fig1_221831959.

https://www.sciencedirect.com/science/article/pii/S1532046417301764.

Conclusion

As underlined by recent positions taken by clinicians on the clinical implementation of predictive AI models[17], it seems necessary to adjust procedures concerning artificial intelligence algorithms where there is a risk of non-transposability of the model, along with guaranteeing their interoperability with clinical IT systems. The learning system for rare diseases, based on the search for similar patients (figure 1), responds to these demands as well as to the requirements for explicability formulated by clinicians. This learning system considers all the information available for each patient whether generated during care or research, and integrates all its various forms. If algorithms must be shared, the complexity of the data requires local governance and a federated approach. A centralized approach risks undermining the confidence of patients and clinicians alike, and above all, deprives us of the wealth of raw data that is indispensable for producing algorithms with high clinical added value.

17 Nong P. and al (2024). "How Academic Medical Centers Govern AI Prediction Tools in the Context of Uncertainty and Evolving Regulation", *NEJM AI*, vol. 1–3, online: <https://ai.nejm.org/doi/full/10.1056/AIp2300048>.

Citizens and the second life of health data: the challenges in terms of information and rights

Caroline Guillot
Director of citizen Engagement, Health Data Hub

Stéphanie Combes
Director, Health Data Hub

Every time a patient interacts with the healthcare system, personal data is generated. This data combines precise information about a person with information about their state of health. Initially, this data is used to optimize patient care.1 However, this data can also be used for research: this is the secondary use of health data, its "second life"![2] By analyzing them together, we can learn lessons that will help us to provide better care for all our patients - lessons that will benefit us all.

France has been a forerunner in the secondary use of healthcare data, with a number of initiatives over the past 20 years: the creation of Sniiram in 1999, expanded in 2016 to become the Système National des Données de Santé ("SNDS"),[3] to which the Assurance Maladie has opened up further access and wider availability. The law then broadened the scope of health data that can be shared, including data from certain cohorts, registries, and warehouses, while encouraging artificial intelligence (AI) projects that can make use of the services offered by the Health Data Hub (HDH),[4] created for this purpose. In this way, the Assurance Maladie and the HDH, in conjunction with other players in the healthcare ecosystem and their partners, have worked hard to lay the foundations for a greater use of our data assets.

The challenges associated with the secondary use of health data are numerous: reducing delays in accessing health data, formulating a medium- to long-term vision for national health data assets and the rules of the game for sharing them, supporting local initiatives (hospital health data warehouses or city data warehouses), opening up data, etc. There are also challenges at the European level, with the European Health Data Space (EHDS). Another one of these challenges is to get civil society on board, in particular by enabling it to be informed of projects carried out with health data as well as enabling it to assert its rights. This article proposes to explain this challenge, particularly from the angle of information and rights. It will show what has already been done and what remains to be done to address the challenges in this specific field.

---

1 For information useful for medical monitoring - this is referred to as the primary use of health data. For health data - gathered from multiple sources, cross-referenced or transformed in anticipation of subsequent analysis, whether or not it uses the most innovative techniques available (such as artificial intelligence) – we refer to this as the secondary use of health data.

2 Menager K. et *al.* (2023). *Qualitative study to assess citizens' perception of sharing health data for secondary use and recommendations on how to engage citizens in the EHDS, Deliverable 8.1*, Tehdas, online: < https://tehdas.eu/tehdas1/app/uploads/2023/03/tehdas-study-to-assess-citizens-perception-of-sharing-health-data-for-secondary-use.pdf>.

3 Created by the *Law of January 26, 2016 for the modernization of the French healthcare system*, the SNDS gathers the main existing public health databases. The purpose of the SNDS is to make these data available to promote studies, research, or evaluations of public interest that contribute to one of the following purposes: informing about health; implementing health policies; understanding health expenditures; informing professionals and establishments about their activities; innovation in the fields of health and social care; surveillance, monitoring and public health security.

4 Created by the *Law of July 24, 2019* on *the organization and transformation of the healthcare system*, the HDH is a public structure giving players easy access to non-nominative healthcare data, hosted on a secure platform. Project leaders can cross-reference data and analyze it to improve the quality of care and support for patients.

The challenge of better documenting the benefits of public utility projects

When it comes to the secondary re-use of healthcare data, France has considerable assets to offer: from the large-scale medico-administrative database of the Assurance Maladie, to the vast research cohorts, as well as the current momentum in the deployment of hospital healthcare data warehouses, all of which form part of the SNDS as defined by law. The exploitation of this wealth of data is extremely promising.

Indeed, the secondary use of data is the source of multiple uses that were also highlighted by the health crisis. A recent article[5] sums them up as follows: "the benefits expected from the accessibility of this data:

1. They facilitate the epidemiological monitoring of major diseases, as this is the basis for many public health decisions;
2. they make it possible to measure the effectiveness of care and detect local or regional situations that deviate significantly from the average;
3. to monitor changes in practices over time and their impact;
4. they enable us to document the positive and negative effects of innovations introduced as a result of changes in practices, be they medicinal, technical or organizational;
5. they provide information on patient access to healthcare interventions: rates and delays can be calculated and cross-referenced with geographical variables, social data, and economic data, so as to identify inequalities, what works and what doesn't, and where to prioritize and focus efforts to improve;
6. they contribute directly to improving knowledge, which is an essential step in improving practices as well as designing effective and protective public policies;
7. they are indispensable for developing new uses of artificial intelligence through high-quality algorithms because they would be trained and tested on representative populations, which is not always the case at the moment;
8. they make it possible to assess the effectiveness of prevention or screening campaigns, as well as compliance with best-practice recommendations."

To illustrate projects that use health data, let's take the example of one project that the HDH is supporting.[6] The *Jumeau Numérique "Etat de santé d'un territoire"* project, supported by the VYV Group and Datalab Normandie, aims to improve the prevention of pollen allergies using health and environmental data. In France, 4 million people are diagnosed with asthma and 12 million suffer from allergic rhinitis. Respiratory allergies are the leading chronic ailment among children. They are on the increase, affecting patients' daily lives with recurrent episodes of wheezing, shortness of breath, chest tightness, and coughing. It usually takes seven years from the first allergic symptom to diagnosis: a diagnostic and therapeutic lag that exacerbates the disease and the risk of developing related pathologies.

This project combines health and environmental data with a view to linking drug prescriptions, pollen episodes from the Réseau National de Surveillance Aérobiologique (RNSA)[7] (pollen types, volumes, etc.), and local characteristics: plantations, trees, crops, weather. The database is anonymized so that it can be made more accessible via data storage and retrieval tools. More specifically, interactive 2D and 3D maps are provided for each commune, as well as projections linked to weather forecasts.

5 Aymé S.and *Al.* (2023). *Bénéfices et risques de l'utilisation des données de santé à des fins de recherche : Rapport du Conseil scientifique Consultatif du Health Data Hub*, online : < https://hal.science/hal-04345572v1>.

6 The "Jumeau Numérique" project was the winner of the Datalab Normandie call for projects launched in 2021, and it was supported by the HDB. Within this framework, the HDB's contribution consisted in providing anonymized healthcare data: enumeration of patients who had at least one dispensing of allergen-related drugs between 2010 and 2019 by department, gender, and age; ranking of communes in the region within which at least one dispensing was made.

7 The RNSA is an association whose main objective is to study the content of biological particles in the air that may have an impact on the allergic risk for the population.

In this way, the project reveals new data on the location and distribution of allergens and offers visualization and prevention tools for health problems linked specifically to air quality. The case study created was presented in the cities of Caen and Rouen. The analyses carried out over eight years and the forecast simulations are both promising for improving the generation of pollen alerts for the general public, as well as for better integrating this type of factor into regional planning.

Several hundred projects a year are carried out using collective data assets. This data is accessible to all within a strict regulatory framework. Healthcare establishments, start-ups and manufacturers, associations of healthcare system users - all players can and do have the legitimacy to use the data for purposes of public utility. These projects are the work of teams of people with a wide range of skills. Some of them benefit from support from the Health Data Hub, notably in terms of legal assistance, expertise in data use, access to a state-of-the-art technological platform for analysis, access to basic numerical indicators, etc.

To bring these projects to fruition, the teams gradually reach milestones such as data matching, user testing of the applications designed, and so on. Users are kept informed of these projects, above all because of the transparency obligation incumbent on those responsible for handing the data. This transparency is a direct result of the application of the General Data Protection Regulation (GDPR), which makes it an obligation and sets out the guidelines in Articles 12, 13, and 14. In particular, it consists in making available information on research projects and on people's rights with regard to data in a clear and easily accessible form. At the SNDS, data is referenced in a directory that is accessible to the public. To date, citizens have access to a wide range of information on projects that use their data in a de-identified way, including the regulatory framework for data processing, its objectives and its interests for public health, etc. They also have access to the results, often in the form of a scientific publication.

However, the actualization of these benefits remains relatively under-researched and under-reported. Only certain facts have been given particular visibility, such as the use of Assurance Maladie data during the Mediator crisis,[8] or more recently the Epiphare studies during the health crisis. Yet these examples where health data have played a decisive role for the healthcare system are essential, not only because they will make citizens more aware of the role of data, but also because they are essential in providing an assessment of the benefits linked to the use of said data, enabling informed decision-making by those responsible for formulating public policy relating to health data. More work needs to be done to identify a series of indicators to describe the concrete benefits of using health data for users of the healthcare system. This may depend on the categories of use that need to be proposed. Typically, in the case of the Mediator affair, health data made it possible to highlight a side effect, and so the benefit can be estimated by the number of people impacted by this discovery. The benefit could also be that of having made people aware of the value of the secondary use of health data. The challenge is to highlight the benefits which are actually observed for the healthcare system from the secondary use of data, which could become a means of assessing whether databases are being used optimally.actually observed for the healthcare system from the secondary use of data, which could become a means of assessing whether databases are being used optimally.

8 For more information on the mediator crisis, see: Radio France Internationale (2023). "French pharmaceutical company appeals conviction over weight-loss drug", online: <https://www.rfi.fr/en/france/20230109-french-pharmaceutical-company-appeals-conviction-over-weight-loss-drug>.

The challenge of making room for dialogue with civil society

As mentioned above, the effort to disseminate information can consist of showing the benefits of projects by clearly explaining how they meet a public health goal, publishing the results obtained from these projects (which, moreover, is a mandatory condition for the SNDS), using different tools such as awareness-raising videos[9], and building indicators. But more generally, it seems necessary to build a culture of health data and disseminate it to enable people to become fully aware of the benefits of using data for research and, ultimately, to better understand their rights relating to health data and exercise them in an informed way. This is even more important given that in 2022, according to the HDH and France Assos Santé (FAS) health data knowledge barometer, only half of people in France said they were aware of their rights concerning their personal data, and only one in five knew them in detail. What's more, 77% of those surveyed felt they were insufficiently informed about research projects using health data, and a majority expressed a desire to learn more, particularly about the use of data for research.

In this respect, a new tool is currently being developed by the HDH to facilitate the exercise of citizens' rights over their health data. It combines a digital form for exercising rights with an educational portal providing information on health data and research. This innovative tool responds to the informational needs of players in the healthcare ecosystem and civil society concerning health data. In 2023, thanks specifically to feedback from a working group, a first version of the patient portal and the national rights management form will be put into production, making it easier for those concerned to exercise their rights. Two committees have been set up and have been meeting since January 2023. The steering committee, made up of representatives from FAS, the DREES (Direction de la recherche, des études, de l'évaluation des statistiques),[10] the CNIL (Commission Nationale de l'informatique et des libertés)[11] and the HDH, and the citizen committee, made up of representatives of healthcare system users and student associations.

The committees agreed on the importance of developing educational information, with particular emphasis on the fact that health data can be used for research. By raising public awareness in this way, rights will be better understood and citizens will be able to assert them if they so wish. Other questions considered as priorities by the committees that met recently include: what rules must be respected in the use of this data? What data is targeted by this tool, and what rights are associated with it? What projects are currently underway? All of these questions will be addressed in an accessible and educational way in a citizen information portal. Its aim is to provide all the answers you need to make informed decisions when exercising your rights or when simply looking for information. The aim is to prevent people from making choices about exercising their rights by default or through lack of information.

Users of this information portal can be directed to a dedicated space for asserting their rights. This will consist of a form enabling a citizen to request the exercise of their rights relating to personal data. This form will be sent to the Data Protection Officer (DPO) of the relevant database. A secure technical procedure will enable the rights requested in the form to be asserted. An email will be sent to the requester to inform them that this has been done.

9 Ford E.and *Al.* (2019). 'Our data, our society, our health: a vision for inclusive and transparent health data science in the United Kingdom and beyond', *Learning Health Systems,* online: <https://doi.org/10.1002/lrh2.10191>.

10 The Directorate of Research, Studies, Evaluation, and Statistics (DREES) is the French ministerial statistical service in the fields of health and social affairs.

11 Created by the law "*Informatique et Libertés*" of January 6, 1978, the CNIL is the French authority responsible for ensuring the protection of personal data contained in computerized or paper files and processing, both public and private.

In addition to the benefits of enabling people to read information for which they would feel concerned as a result of this health data education, the construction of a health data culture, as well as the introduction of more pedagogy, would most certainly contribute to deconstructing the image of apprehension, or even citizen reluctance[12] to share, that remains entrenched among other stakeholders.[13] The aim is therefore to encourage a societal movement to acquire a "health data culture", by accelerating the creation of a framework of thought that would give meaning to the very idea of the secondary use of health data, notably so that an educational action concerning data can have concrete effects.[14] This raises the question of how to include players in the health data governance system, as well as the types of collective players capable of providing a different representation. Users and their patient associations are excellent vectors for bringing a digital solution to the attention of the community. Thanks to the development of specific communications networks, they represent an additional opportunity for visibility for the digital tool. They can encourage other users to use the system by sharing their positive experiences.

Finally, creating a dialogue around this subject, and more broadly about AI, with millions of French citizens constitutes another response. This is also outlined in one of the proposals in the report by the Commission on Artificial Intelligence. In fact, this is one of the proposals made in the report by the Commission on Artificial Intelligence.[15] It indicates that we will not be able to move forward collectively and build a unified system if we do not have a common project, a shared narrative – a project of collective progress; and that it is therefore necessary to engage in debate. This openness and dialogue can take a variety of forms, including studies (on citizens' perceptions, needs, expectations and knowledge of health data),[16] participatory and co-design workshops (based on target or broader audiences, to raise awareness of health data, provide training, and convey citizens' commitments to data), consultations and collaborations. Workshops for scientific popularization, transparency initiatives, open data projects, and citizen cafes can also be implemented.

Conclusion

The issue of informing citizens about the use of their data, particularly for research purposes, is a multidisciplinary one, as lawyers, communicators, managers of healthcare structures and warehouses, associations, and professionals collecting data are all involved. One of the challenges of the secondary use of healthcare data is to enable civil society to play an active role in understanding the data. The challenge is to provide information that is both educational and participatory, a challenge that concerns all the organizations that use or make data available. Two approaches seem to address this: the approach of finding the right balance between legal and educational information; and the participatory approach of strengthening dialogue.

12 État d'esprit Statis and Health Data Hub (2023). *Conférence de consensus : Comment sensibiliser la société au partage des données de santé, Rapport d'étude*, 2023, online : < https://www.health-data-hub.fr/actualites/conference-de-consensus-les-propositions-citoyennes-pour-parler-du-benefice-des-donnees>.

13 Neves A. L. and *Al.* (2019). 'Health Care Professionals' Perspectives on the Secondary Use of Health Records to Improve Quality and Safety of Care in England: Qualitative Study', *Journal of Medical Internet Research* 21(9), e14135, online: <https://www.jmir.org/2019/9/e14135>.

14 Combes S. et Guillot C. (2021). « Utilisation secondaire des données de santé : quelles modalités de participation citoyenne », Third, n°7, online : < https://third.digital/numero-sept/utilisation-secondaire-des-donnees-de-sante-quelles-modalites-de-participation-citoyenne/>.

15 Commission de l'intelligence artificielle (2024). *IA, notre ambition pour la France*, online : <https://www.info.gouv.fr/upload/media/content/0001/09/4d3cc456dd2f5b9d79ee75feea63b47f10d75158.pdf>.

16 Guillot C., SCHWEIGHOFER A. and LUCAS J. (2024) « Comment parler des données de santé aux citoyens ? Le défi de l'information créative et participative des structures de santé » dans Bubien Y. and Bataille-Ambert A., dir, *Innovations et communication en santé*, Éditions LEH.

obvia

obvia.ca