SMART CITIES AND INNOVATIVE USES FOR PERSONAL DATA: scenarios for using data to restore the balance between public and private spheres

Geoffrey Delcroix
Innovation & Foresight Project Manager,
Department of Technologies and Innovation - CNIL

CAHIERS IP
INNOVATION & PROSPECTIVE
N°05
La plateforme d'une ville
Les données personnelles au cœur de la fabrique de la smart city

Geoffrey Delcroix manages innovation, research and foresight projects at the Department for Technology and Innovation of the French Data Protection Authority (CNIL). A graduate in political sciences, geopolitics and defense, Geoffrey Delcroix began his career as a consultant and researcher in the Futuribles team, an independent center for contemporary world study. He then headed up the foresight unit at the French Ministry of the Interior's foresight and strategy team before joining CNIL in 2011.

The team focuses on three missions:

- explore emerging trends at the frontier between digital technologies, ethical issues and data
- exchange ideas and act as the main point of contact for innovation ecosystems (the team works with startups, labs and academics around those topics)
- experiment with innovation methods and produce or co-produce demos, proof of concepts and prototypes relating to privacy issues.

The team publishes on various topics (connected vehicles, chatbots, robotics, AI, connected objects, drones, digital health, algorithms, etc.). All articles are available from LINC (https://linc.cnil.fr/), the platform for CNIL's innovation and future-focused media.

The Platform of a City, the fifth IP Report, is an exploration of the issues related to smart cities and data uses in urban planning and services. It contains recommendations, in particular regarding the different tools that can be used in the future to create meaningful and controlled uses of personal data for general interest purposes.

KEYWORDS

- OPEN DATA
- PERSONAL DATA
- INFORMATION COMMONS
- FREE FLOW OF DATA
- PUBLIC-PRIVATE PARTNERSHIP

In the face of the contradictory imperatives of the smart city — personalizing everything while respecting the right to privacy, optimizing without rejecting — and in response to the new landscape, particularly the arrival of major data companies, the challenge now is to produce new models for regulating city data, ones that respect individuals and their freedoms.

INTRODUCTION

How should data that offers powerful added value for the general interest, but is collected and used by private actors, be shared with public actors while respecting the rights of the businesses that collect and process the data as well as the rights and freedoms of the individuals concerned? This is the question that laws and public policies are currently trying to answer. Other sections in La plateforme d'une ville *(The Platform of a City, available online in French only[1]), published by the Innovation and Foresight unit at CNIL, the French data protection authority, describe how the digital city's new services rely increasingly on personal data that is collected and processed for commercial ends by private actors.*

This data, which does not fall within the natural ambit of a public service (whether directly managed, under concession, etc.), does nonetheless interact profoundly with issues of public service and can be invaluable in the delivery of a public service mission.

At present, a number of different tools are being developed by the various stakeholders in this debate. All these tools have serious limitations but also represent real opportunities. Each relies on achieving the right balance of rights and obligations between the various actors involved.

These tools can be characterized according to two features. First are the legal obligations they impose on private actors: among the four proposals described below, some could be rolled out within existing legislative frameworks whereas others would require new legislation before they could be put into practice. Then

[1] https://linc.cnil.fr/la-plateforme-dune-ville-explore-les-enjeux-de-la-smart-city

comes the question of data granularity: in some cases very fine data, including personal data, is sent to the public actor; in others, the public actor can access data only once it is aggregated and anonymized.

In a previous report, Partage ![2], we showed how a traditional regulatory model in isolation has little chance of being effective, and that a regulatory posture adapted to these platforms requires a new and more dynamic balance that would employ a palette of regulatory mechanisms, which would provide a range of levers to impact: the balance of power between actors (market); technical systems and architecture (technology and design); ground rules (regulator and standards); and, lastly, self-determination and returning power to the individual (empowerment).

By combining the two features (legal obligations and data aggregation) with the four regulatory levers, we obtain a matrix of four distinct scenarios that represent as many possible futures, as alternatives or in combinations, for new forms of data sharing.

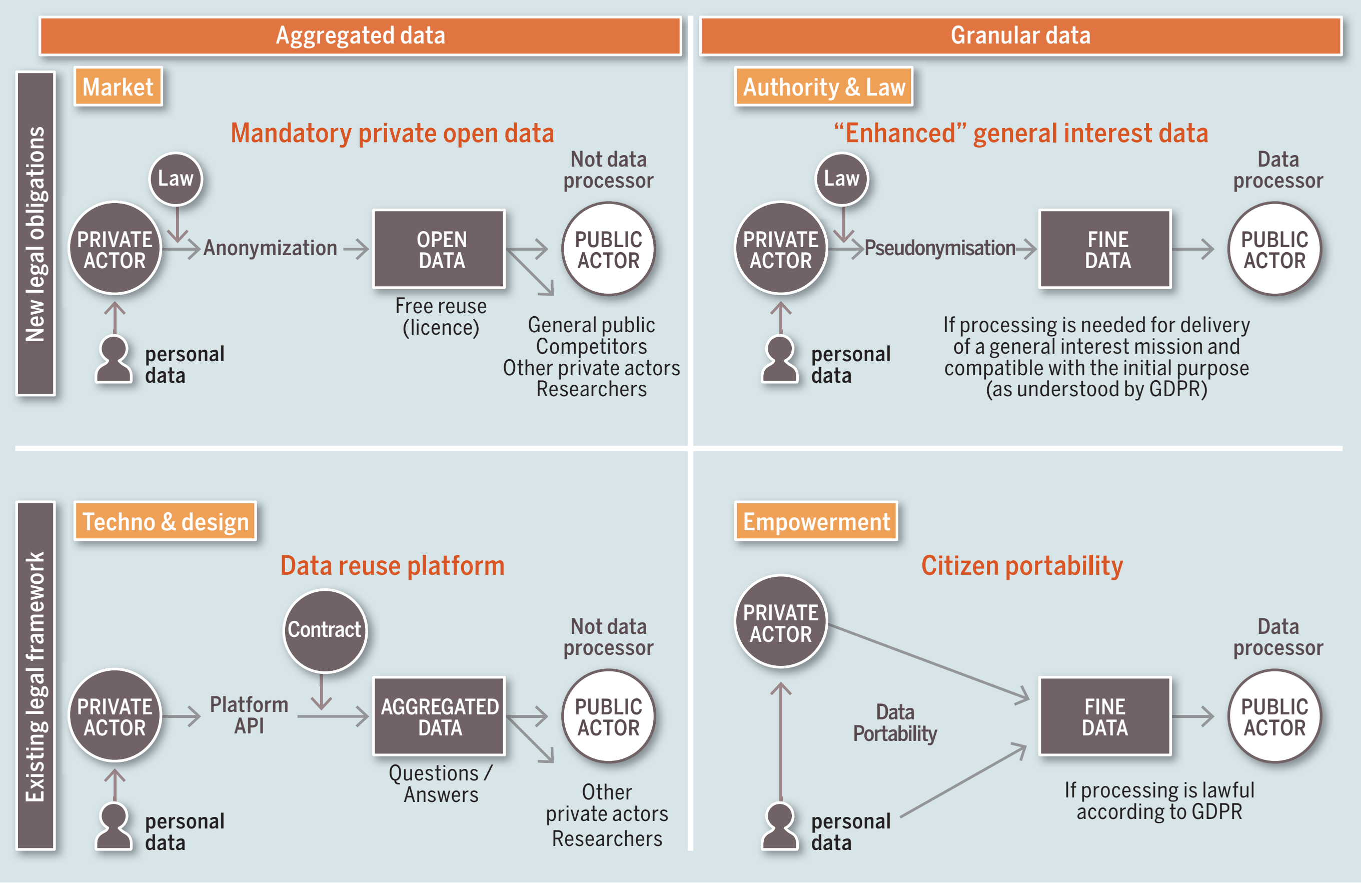
These scenarios offer different options for dividing the challenge of exploiting fine data and reassigning the capacity to take action in the general interest, redefining the balance of power between public and private actors within the realm of public service.

They differ in how they allocate responsibility for personal data protection, which can lie with either the private or the public actor. Whatever the scenario, the challenge is to establish best practices guaranteeing that the rights and freedoms of the people providing data are respected.

Without making any judgments about one or other of these mechanisms, setting out the basic structure of each and highlighting their potentialities serves to identify the questions raised in terms of protecting people's personal data.

[2] See the Innovation & Foresight report *Partage ! Motivations et contreparties au partage de soi dans la société numérique* (Share! Motivations and counterparties to sharing the self in the digital society). In French only
https://linc.cnil.fr/fr/dossier-partage

A tentative matrix of possible futures for data sharing



GENERALIZING OPEN DATA FOR THE PRIVATE SECTOR

Acting on balances of power and creating the conditions for effective self-regulation might involve setting up mandatory private sector open data policies for data with proven impact on the efficient operation of the market or policies in the general interest.

Private actors would be under a legal obligation to provide open data access to certain data they hold, for example as provided for under two French laws, the so-called "Macron Act" and the Act for Energy Transition.[3] In order for this process to meet personal data protection requirements, in most cases implementation would involve anonymization processes that comply with certification requirements.[4]

The advantage of this mechanism is that data can be reused without restriction, by competitors, public bodies, researchers, citizens, etc. However, it is not without its drawbacks. Anonymization comes at a price: financial for the private actor and in terms of the loss of dataset information for other users; public actors will not have access, for instance, to the very fine data that contributes to the success of general interest missions. Private actors remain in control of the quality of the datasets retrieved.

EXTENDING GENERAL INTEREST DATA BEYOND PUBLIC SERVICE CONCESSIONS

To change the ground rules is to take the view that overriding higher interests justify delineating the intangible limits society has set on ethical and political subjects. In this scenario, the issue is allowing and regulating reuse of personal data by public actors for certain purposes in the general interest, but without infringing the rights of individuals. This would involve extending the scope and modalities of the emerging notion of general interest data. Currently, general interest data is limited in France to companies operating public service concessions. It would in this scenario be extended to private actors, with the exclusion of their contract relations with public authorities.

OPEN
Le restaurant que vous suivez sur Instagram est à 200m
LAST STOP
Times Sq-42 St Subway Station
La Statue de la Liberté ferme à 19h.
OPEN

CNIL – Five BY Five – ©Léa Chassagne

This data is currently anonymized by the private actor prior to being made available as open data. The idea would be to open the way for certain fine data to be provided to public actors for public service missions; the public actors would then be responsible for data anonymization where it is made available as open data.

A balance of interests should make it possible to avoid harming the interests of a private actor that had invested in proprietary data processing and also to avoid violating individuals' right to privacy, as they would have consented to data processing within the context of a specific service. Public authorities become responsible for data processing and must respect all applicable rules (legal basis, purpose limitation, compliance to all data protection principles, etc.).

Such a mechanism would offer the advantage of resetting the balance of powers between certain private actors and public authorities, which would form an effective lever for successfully accomplishing general interest missions without any infringement of the rights of individuals. The drawback with this scenario is its burdensome nature: for private businesses obliged to restitute data and for public organization users responsible for personal data protection.

This scenario has a number of backers. In the wake of France's Act for a Digital Republic, which set out the broad lines, and the 2015 report on general interest data by the French Ministry of Economy[5], similar hypotheses have been developed by the European Commission in its work on the free flow of data[6] and in Luc Belot's report to the French parliament[7], which calls for the definition and identification of a "territorial interest data" category.

[3] Act 2015-990, August 6, 2015, for Growth, Activity and Equal Economic Opportunities, and Act 2015-992, August 17, 2015, for Energy Transition for Green Growth

[4] See Article 29, Working Party (European Union) 05/2014 opinion on Anonymization Techniques

[5] CGEIET and IGF. Report on general interest data, September 2015. https://www.economie.gouv.fr/files/files/PDF/DIG-Rapport-final2015-09.pdf

[6] Commission staff working document on the free flow of data and emerging issues of the European data economy. Accompanying a document on Building a European Data Economy, January 2017

[7] Luc Belot. *De la smart city au territoire d'intelligence[s]*. Report to the prime minister on the future of smart cities, April 2017

PERMITTED REUSE UNDER CONTROL OF PRIVATE ACTORS

In regulatory terms, acting on systems and architectures is simply to keep step with current technical transformations in the data economy. This might involve using legal and technical measures to regulate the emergence of platforms for data access and sharing. Responses to open data, data lakes and mass anonymization might follow the API model, with data taps and differential privacy.

Private actors could use tools such as APIs etc., to set up a platform for reusing their data that would enable the reuser to exploit some data without actually processing it: the reuser would interrogate a database held by the private actor and receive only the answer, not the full dataset. Properly designed, such a system would enable rich data exploitation while minimizing the risks of infringing individual rights. In addition to anonymization, the platform could deploy two further types of tools:

- legal: a contract must govern what reusers may or may not do; for example, a clause prohibiting a partner from attempting to reidentify people and thereby compromise their anonymity, as well as clauses detailing how liability is to be apportioned;
- technical: real-time audits, controls, checks and log analyses to deliver dynamic risk analysis, for example to limit the chance of database inference attacks.

The advantage to private actors of such a mechanism, which would not need new legal obligations, is that they would not be required to open up their data en masse and would not have to bear responsibility for personal data protection. The drawback of this scenario is the cost to private actors of developing and maintaining a platform, although this could also offer new opportunities and revenue streams via the sale of anonymized data.

ENACTING CITIZEN PORTABILITY

The new regulations governing personal data protection offer everybody the opportunity to determine how their data is used and empower citizens to participate in missions of general interest.

The General Data Protection Regulation (GDPR) introduces a right to data portability that promotes the reuse of personal data by a new processor, without any obstruction by the initial processor, and under the exclusive control of the person concerned. This arrangement, which will enable users to migrate from one ecosystem of services to another (competing or not) bringing with them their own data might also enable them to opt in to citizen portability to benefit general interest missions.

Communities of users could exercise their portability rights in relation to a service in order to provide a public actor with access to their data for a specific purpose relating to a public service mission. The public actor is then responsible for data processing and is therefore also required to respect the principles of data protection.

Such a mechanism would have the advantage of creating new datasets for use in public service but without imposing new legal restrictions on private actors. The drawback for this scenario is the critical mass required as widespread acceptance and participation will be needed to constitute relevant datasets. The incorporation of simplified, innovative and non-restrictive opt-in systems should help ramp up participation levels.

> ***"IN A MORE FUTURE-FORWARD VISION, A PROCESS SUCH AS THIS COULD LEAD TO BOTTOM-UP CREATION OF AN INFORMATION COMMONS, BUILT BY INDIVIDUALS IN THE GENERAL INTEREST. THIS WOULD ENTAIL BUILDING GOVERNANCE PROCESSES FOR THE INFORMATION COMMONS, PERHAPS IN THE FORM OF PUBLICALLY OWNED AND MANAGED LOCAL DATA CORPORATIONS."***

In a more future-forward vision, a process such as this could lead to bottom-up creation of an information commons, built by individuals in the general interest. This would entail building governance processes for the information commons, perhaps in the form of publically owned and managed local data corporations.

CNIL takes the view that any adjustments to the balance of power between private and public actors concerning city management and intended to improve public policy must go hand in hand with greater oversight of public authorities. They will be required to adhere to GDPR[8] and specifically the notion of legitimate purpose in regard to reuse of the data provided.

REGULATION THROUGH THE COMMONS AND A DEDICATED GOVERNANCE STRUCTURE

In the face of the contradictory imperatives of the smart city—personalizing everything while respecting the right to privacy, optimizing without rejecting—and in response to the new landscape, particularly the arrival of major data companies, the challenge now is to produce new models for regulating city data, ones that respect individuals and their freedoms.

Innovative and efficient regulatory methods are an interesting area, for example commons-based production and governance of city data, with the establishment of new governance structures for

[8] General Data Protection Regulation (GDPR) is the commonly used name for the European Union legal framework, adopted in 2016, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data: http://data.europa.eu/eli/reg/2016/679/oj

this data. Adoption of these types of mechanisms would also deliver valuable tools for aligning with the European GDPR, for instance in terms of the core notion of consent.

DEFINING THE COMMONS

In 2014 Valérie Peugeot addressed the question of data in the smart city from the commons perspective, suggesting "moving beyond the strict boundaries of personal data to examine digital data as a whole [...] inspired by Elinor Ostrom's work [...] to develop a data sphere for the commons, by which is meant data that can be categorized as a collective resource, and that falls neither within the property regime managed by public authorities in the narrow sense, nor within the market system." The commons system relies on management of the relevant resource by a community, which structures its governance rules around what is termed a bundle of rights. Valérie Peugeot suggests extending the commons to data in the public sphere, data produced under share-alike licenses (Wikipedia, OpenStreetMap, etc.) and certain data produced by private businesses. To go further in the direction of commons-based production, it will probably ultimately have to include open data reference data and general interest data as defined in the Act for a Digital Republic and other public interest data as may be defined by future legislation. We could cite as an example data held by major data companies such as Waze, collected from users on a data-for-services basis.

These businesses, which claim to work in the general interest, would then cease to restrict the general interest to the sum of their clients' private interests, allowing public authorities access to reuse the data that they themselves exploit. The recommendations outlined above (extending the notion of the general interest data and activating citizen portability) could be of use in enabling this approach to develop.

This approach, based on the commons and moving beyond the open data mindset, has been gathering strength in recent years. CNNum (the French Digital Council), in an opinion issued in April 2017 on the free flow of data in the European Union, suggests mechanisms for data sharing[9]: "Member States could encourage different players to share their data on a voluntary basis in order to contribute to a research program, an industrial project or a public policy, either occasionally or on a long-term basis. The pooled data could be collected by a public body and be aggregated before being reused or redistributed [...]." Regarding general interest data held by the private sector, the report suggests invoking a general interest motive to require the data to be handed over, notably for the purposes of managing public sectoral policies, providing information to citizens, and economic development. There is no infringement on property rights where data is handed over only to public authorities or is reused for non-commercial purposes. In cases of reuse for commercial purposes, the report states that indemnity payments are the only solution that avoids structurally undermining private actors. And here we have one of the key challenges of the commons-based approach, which is currently relatively conceptual: there is undoubtedly an interest for the sum of all parties, but the gain for actors who are currently in a position of strength in terms of data is more uncertain. The aim is therefore to find a way to maximize value to society as a whole but without disincentivizing the actors responsible for creating this new data.

GOVERNING THE COMMONS TO BETTER PROTECT PERSONAL DATA

Commons for the city cannot be constituted without establishing modes of data governance. In its opinion, CNNum gives the sectoral example of the US Bureau of Transportation Statistics, which aggregates air traffic data from US airlines. But others go further, suggesting trusted third party actors for a given territory, which would offer shared governance tools able to enforce compliance, particularly with personal data protection rules. This is the type of model put forward by Datact in the form of its publically owned and managed local data corporations (*Régies de données*)[10], third-party legal entities with governance shared between the city as public actor and its various stakeholders—a true commons of the city but also a data interrogation and processing system allowing data flows to the various actors requiring them to be opened and closed on demand. This third-party actor would facilitate data flows between the various stakeholders, acting as a hub and monitoring the admissibility of data processing purposes. It would also ensure that applicable licenses were respected and personal data protected by providing mechanisms for registering consent.

Such an arrangement would also make it possible to move beyond the mindset of automatic anonymization of city data. It would be possible, as proposed by the Open Algorithms project[11], to allow certain actors to use data without possessing it and in full compliance with the rights of individuals. A management tool of this kind would offer the advantage of opening up city data and resetting the balance of power between the public actor and private actors not bound by contracts to the public sector. It would provide interested small businesses, collectives, residents and non-profits with a means to reappropriate ownership of the city commons, and most importantly it would allow data reusers that wished to process personal data to ask for explicit informed consent from the individuals concerned.

[9] CNNum, Opinion of the French Digital Council on the Free Flow of Data in the European Union, April 2017, (in English) https://cnnumerique.fr/wp-content/uploads/2017/05/OpinionCNNum_FFoD_ENG-1.pdf

[10] *Concevoir une régie de données territoriales - Vers une nouvelle fabrique de services urbains*, (Designing publically owned and managed local data corporations - Towards new methods for imagining city services), dossier produced by Le hub, Chronos and Datact, in *La gazette des communes*, May 2014

[11] http://www.opalproject.org