

How to define and regulate “data of general interest”?

Bertrand Pailhès

Abstract:

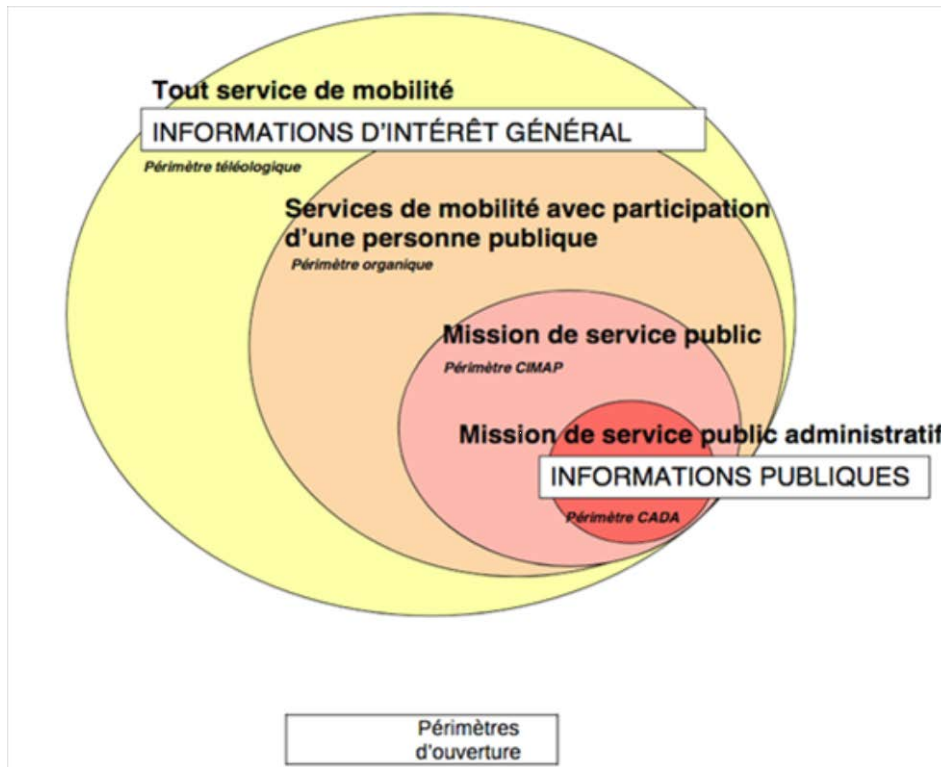
Since 2015, several reports have been made about opening data (private and public) “of general interest” for the sake of stimulating innovation and curbing the power of certain players in the market. This “opening” fits in with a vision of “shared” digital resources, of data as the new window of opportunity (following scientific knowledge and cultural goods). After distinguishing “data of general interest” from “open data” and from the usual means of access by public authorities to private data, attention is turned to the grounds underlying this trend and the obstacles in its way. A grid is proposed for analyzing the means for protecting the interests of the parties involved while allowing data to circulate and be shared for the common interest.

In 2015, a fierce debate about transportation data broke out in France between public transit firms (RATP, SNCF, Transdev, etc.), the newcomers in this market from the digital economy (Google, Uber and, too, Citymapper) and public authorities, who wanted to foster the opening of public data.¹ Francis Jutland submitted a report in March 2015 that cleared the way for the legislative measures incorporated in the Macron Act.² This report defined “*information of general interest*” as any “*data (or classes of data) from services of mobility for which it is deemed to be the right time to open them*” (JUTLAND 2015, p.72). This covered data from public transit services and, too, from private transportation firms. Although the act of law concentrated on transporters who are granted a “mission of public service”, this was the first report in France to broaden discussions about the opening of public data and provide a full view of the issues related to “open data” in a sector (transportation) where public and private organizations play essential roles.

¹ This article, including quotations from French sources, has been translated from French by Noal Mellott (Omaha Beach, France). The translation into English has, with the editor’s approval, completed a few bibliographical references.

² Act n° 2015-990 of 6 August 2015 for growth, activity and equal economic opportunity available at <https://www.legifrance.gouv.fr/eli/loi/2015/8/6/EINX1426821L/jo/texte>.

Figure 1: Information of general interest in transportation services
 Source: JUTLAND 2015, p.72.



In September, a report on “*data of general interest*” (CYTERMANN *et al.* 2015) expanded the legal and economic framework of this discussion to other fields. It identified the levers for, and obstacles to, the access to such data. The press release on the data economy issued by the European Commission in January 2017 also raised the question of access to data for public and private organizations. The “sharing” of public and private data also figured at the center of the report on a French and European strategy for “*giving meaning to artificial intelligence*” (VILLANI 2018).

Data of general interest for what purpose?

The example of transportation provides the best illustration of how data are to be collected and shared among private and public organizations. In this branch of the economy:

- Data are abundant.
- We immediately understand that massively collecting them would be useful to cope with everyday problems, such as traffic jams, air pollution and the operation of public transit.
- The data are collected by a wide variety of organizations, ranging from the most private (the individual car) to the highly public (an infrastructure in the public domain, such as a subway system).
- The transportation sector is drawing newcomers who are perceived as being powerful. In particular, big online platforms have realized that users unanimously appreciate a management of everyday transportation, whether in cars (Waze) or public transit (Citymapper, Google Maps).

There are two stakes in opening data: on the one hand, to develop the economy by boosting innovation and avoiding the formation of situations of economic rent; and on the other hand, to improve the steering of public policy-making.

From the economic viewpoint, data are a “nonrivalrous” good: the use of data by one person does not deprive others of using it. The marginal cost of reproducing data is nearly zero: many users can put the same set of data to many uses; and this maximizes its potential value. Supporting the sharing of data among economic agents is, first of all, a way to boost innovation and make sure that this resource (data) can be put to economically efficient uses.

Furthermore, public authorities have realized that the control of data is important for the economy and society. This control is increasingly a factor shaping the firms and services that will take the lead in the coming years in all branches of the economy. For instance, the determination to develop a European industry of artificial intelligence (VILLANI 2018), which will necessarily rely on huge quantities of data, entails dealing with the question of the access to the data over which firms or administrations have a monopoly. We need but think of the data from users browsing the Internet held by Google or Facebook, or the health data held by French administrations.

From a political viewpoint, the debate about data of general interest refers to a vision of the Internet as a “commons”, *i.e.*, a pool of resources that belong to no one and can be used by everyone under condition that the rules of a shared governance are set. Freeware, Internet protocols and Wikipedia are the most visible examples of this. In the coming years, data will be important for furthering this vision, which is similar to the academic community’s view of scientific knowledge. The objective is to increase transparency and thus exercise control over data-keepers in order to ultimately build up the confidence of users in digital devices.

How to define data of general interest?

Data of general interest are data, public or private, of which the sharing and opening are in the general interest, *i.e.*, a broader interest than that of the party who has the data.

It is essential to make the distinction between “data of general interest” and “open data”. For the latter, lawmakers, when opening public data and restricting as much as possible the conditions of reuse (licences, prices, etc.) without any prior verification of the quality of this reuse, had a twofold objective: transparency and innovation. Open data has as grounds the constitutional rights of citizens to know what the administration is doing. This is not so for data of general interest: the access to them is not a fundamental right and has to be subject to special conditions.

It is also worthwhile distinguishing between data of general interest and the public administration’s access to information in the hands of private or public persons. To collect taxes for example, the administration has to access certain data, while guaranteeing their confidentiality. Data are a new field for applying this prerogative of public authorities.

In the two aforementioned cases, the imperative of the general interest, as set by lawmakers, defines the legal conditions of access to the data. It does not seem necessary to draft new concepts for applying the general interest to these cases.

Data of general interest can be a tool for facilitating the sharing of data in very complex cases, when specific circumstances must be taken into account and under condition that technical, legal or economic obstacles are lifted.

What are the obstacles to opening data of general interest?

To place data-keepers under the obligation to open them to a third party, we must reckon with any legal restrictions that forbid access.

Contrary to a widespread idea, there are currently no proprietary rights on data: intellectual property rights apply to databases *sui generis*, but are not extended to the data themselves. Besides, the data-keeper may invoke secrecy (as in the case of personal data) or security requirements related to the information system to justify limiting the access to data. These legitimate considerations have to be taken into account when weighing the means to be adopted for opening data, means such as application programming interfaces (API, a standardized interface for access to information or features) and security procedures (*e.g.*, for tracing access).

Moreover, rules for opening data might limit the economic freedom of the firm that has the data. Despite its importance, this constitutional principle can, in France, be overridden by the general interest. The question of proportionality is a key factor when defining the conditions for access to data of general interest.

Economically, despite the fact that data are nonrivalrous, the obligation to make them accessible to third parties might reduce the incentives for data-keepers to make investments, since they will be forced to share the data and will not enjoy a monopoly. For this reason, the access to data of general interest ought not to be for free but at a price that sustains incentives for data-keepers to invest or even provides them with an additional source of income. Of course, this price could also depend on the sort of organization requesting access; a public research laboratory does not have a commercial purpose like a start-up.

Opening data of general interest is not to be seen only in relation to a technico-economic environment under the data-keeper's control. Nonetheless, data-keepers have to be able to formulate not only the technical conditions for access to their data (format, interfaces and security measures) but also the conditions for using the retrieved data. In any case, this use is subject to rules of common law (applicable to personal data, regulations specific to business or industries).

How to open data of general interest?

We would start out on the wrong track if we explored the idea of associating intellectual property rights with data. Though attractive owing to the similarity to other immaterial goods (patents, artworks), this idea lets us imagine that the access to data of general interest should be regulated by the economy. Accordingly, the "owner" of such data will open them if he has an economic interest in doing so and in making money from them.

However, just as certain sorts of infrastructure are more efficiently built and kept up by public authorities without private ownership, data that are strategic or indispensable for understanding and solving certain social or economic problems should not be exclusively in the hands of a single agent. The example of cars connected to the Internet illustrates the data jam resulting from this approach. If the automaker "owns" the data for a car, insurers, toll roads, garages and transportation platforms will make arrangements for directly collecting data themselves rather than running the risk that the automaker will set too high a price. It seems, therefore, better for the car's data to be accessible to all these parties under different conditions depending on each party's needs.

The question thus crops up of defining a general legal framework. Can the concept "data of general interest" be legally defined without clinging to considerations specific to the sectors (transportation, housing, the environment) where it is apparently simpler to define the objectives to pursue and the scope of the data concerned? The report on data of general interest (CYTERMANN *et al.* 2015) urged public authorities to adopt this approach with the exception of the access by public offices of statistics to corporate data for the purpose of surveys (Article 20 of

the Digital Republic Act). It also advocated extending this open approach from public administrations to the private firms that provide public services (foreseen in the CADA Act on the access to administrative documents).³ Pursuant to the Macron Act in 2015, the Digital Republic Act of October 2016 continued addressing the issue of the access to certain data in the field of energy or for setting speed limits on highways. Another act has taken part in this trend for opening data by creating a national system of health data.

A general conception of “data of general interest” could be worked out by answering four questions.

- Which data, or type of data, are concerned, and who is the data-keeper?
- Who is asking for access to the data?
- For what purpose is access being requested?
- How will access be provided?

In very few cases is it sufficient to answer only one of these questions.⁴ In general, third-party access to data will be allowed by combining these criteria. For example, a laboratory could justify access to a firm’s data on the grounds that it is a public research organization, but it will have to prove that it is capable of ensuring security. Likewise, a taxi company’s transportation data would *a priori* be deemed of general interest, but the scope and conditions of access would differ depending on the party filing the request for access and on its purpose. The access would thus be different for town hall in Paris and for an online platform that calculates itineraries for drivers.

To deal with this casuistic complexity, a general framework could set two regulatory principles:

- A contract between the parties should override public interventions.
- The custodians of data of general interest should accede to reasonable requests for access. By the way, this proposal is modeled directly on the obligation of access imposed on incumbent operators in the telecommunications sector.

The first principle is evident, since the sharing of data is not a new idea belonging exclusively to public authorities. Many a firm has already had the idea of pooling or sharing data for the purpose of enabling third parties to design new applications or to conduct research. Recently, Transdev set up the platform Catalogue with the objective of pooling data from all transportation operators (from BlablaCar to the SNCF); and Uber has launched Uber Movement to let third parties have access to the data on its taxis. As these examples show, businesses are aware of this issue, and are using data to create an ecosystem with partners who depend on them. However these examples also reveal the need for regulations. Transdev’s initiative did not convince the French railroad company (SNCF) or the Paris subway system (RATP) to provide their data. And Uber alone chooses the data it will open.

Public authorities thus have to assume the role of arbitrator in order for data to be opened to third parties under reasonable (technical, legal and economic) conditions. These conditions should reckon with the four questions already listed, in particular about the party requesting access and their purposes. To increase the possibilities of access, in particular to data in private hands, public authorities should define the criteria for declaring data to be of general interest (and eventually vary the criteria by sector). Among these criteria, let us mention: the importance of data in public policy fields (security, health, environment, etc.), the unique or universal nature of the data (in particular for economic development) or the contribution from users to the data pool (*e.g.*, Waze). Public authorities could launch a transparent procedure (based on these criteria and involving the concerned parties) for declaring types of data to be of general interest and open to requests for access.

³ Respectively French Act n° 2016-1321 of 7 October 2016 for a “digital republic” available at <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746&categorieLien=id>; & Act n° 78-753 of 17 July 1978 on improving the relations between the administration and the public available at: <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000006068643&dateTexte=vig>.

⁴ However that is the case in public statistics (the question on the purpose of opening access), for the judiciary (which has the power to conduct investigations) or for public data (a public administration as data-keeper).

These thoughts, which run through the reports mentioned at the start of this article, should be harmonized at the EU level so as to avoid any “regulation evasion” by the concerned parties.

The need to regulate the access to public and private data of general interest is growing. It seems possible to set up a simple, flexible and general framework that protects the interests of the parties concerned while more widely opening the access to such data and allowing their circulation.

References

CYTERMANN L., DUCHESNE C., VACHEY L., MOREL M. & AUREAU T. (2015) *Rapport relatif aux données d'intérêt général*, September, 93p. Available via:

<https://cdn2.nextinpact.com/medias/report-cytermann.pdf>.

EUROPEAN COMMISSION (2017) “Commission outlines next steps towards a European data economy”, press release, of 10 January, Brussels. Available via:

http://europa.eu/rapid/press-release_IP-17-5_en.htm.

JUTLAND F. (committee chairman) (2015) *Ouverture des données de transport*, a report to the undersecretary in charge of Transportation, the Sea and Fishing in France, March 2015, 147p. Available via: <https://cdn.nextinpact.com/medias/rapportjutand.pdf>.

VILLANI C. (2018) *Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne*, mission parlementaire, March, 242p. Available via:

https://fichiers.acteurspublics.com/redac/pdf/2018/2018-03-28_Rapport-Villani.pdf.