

Economic models of data: A complex relation between supply and demand

Paul Belleflamme,

Professor, Aix-Marseille School of Economics

Abstract:

How are transactions involving data currently organized? From the demand-side, a description is proposed of why and how data acquire value. From the supply-side, questions are raised about where the data come from and who controls their production and collection. Thereafter, the ways that supply and demand meet are discussed. An ever increasing quantity of data is being produced, collected and used; but a small fraction of it is the object of exchanges. Three explanations are proposed related to: the strategic nature of data for firms, the difficulty of organizing decentralized markets, and the lack of control by individuals over the data they produce.

On 19 March 2018, Facebook's stock tumbled following the revelation that Cambridge Analytica had used the personal data of nearly fifty million subscribers without their consent.¹ The same day, we also heard that selfies of cybernauts (eventually used to validate identification procedures) were being sold at a price of up to \$70 on the "dark web".² A few months earlier, the company iRobot took back its previous declarations about reselling the data collected by Roomba, its robotic vacuum cleaner, which make a virtual map of the places cleaned.³ As for Uber, it announced the creation of a platform for sharing for free data on the rides of its drivers and customers with urban planners in the 450 cities where it operates.⁴ These recent events clearly evince how important exchanges of data are in our economies. They also illustrate the various forms of exchanges: freely consented, sharing (Uber), outright theft (selfies on the dark web), or under a contract with more or less clear provisions (Facebook and iRobot).

The intent herein is to help understand how these exchanges of data are organized. For this purpose, I shall start by describing the demand side: why and how do data take on value? I shall then turn to the supply side: where do data come from and who controls their production and collection? In the final section, we shall try to understand how supply and demand meet before turning thought to the possible trends affecting data exchanges in the future.⁵

¹ *The Guardian* and *The New York Times*, respectively: <http://bit.ly/2pIU1sM> and <http://nyti.ms/2u1nLjw>.

² <http://bit.ly/2u8ys3W>

³ <http://bit.ly/2DFoxCv>

⁴ <http://bit.ly/2prPK6B>

⁵ This article has been translated from French by Noal Mellott (Omaha Beach, France).

The demand for data

The demand for data comes from firms and, too, from organizations, such as cities, that are trying to improve their practices. This demand is derivative, since the quest is not for the data as such but for the information extracted from them — ultimately, the knowledge that this information provides and that can be used for decision-making. Thierauf (1999) has defined data as an unstructured collection of facts and figures; information as structured data; and knowledge as information on information. This suggests why the demand for data is a recent phenomenon. The ability to enhance the value of data by turning them into information has increased in recent years owing to the combined effects of “digitization” and “datafication”. Digitization refers to the generalization of electronic formats for storing, duplicating and transmitting data faster and at a much lower cost in terms of energy. Datafication is the multiplication of the digital tracks left by our activities, whether via computers, smartphones, social media or sensors in connected devices. Moreover, a new scientific discipline has developed: data science, which combines tools from mathematics, statistics and computer science in order to optimize the extraction of knowledge from databases.

In summary, the demand for data is expanding because the data available and the tools for processing them are constantly growing. The value of data increases with what has been called the four *v*'s: their volume (whence the phrase “big data” with the implication of economies of scale), their variety (the diversity of their sources with the implication of economies of scope), their velocity (the speed for processing data streams) and, naturally, their veracity (their accuracy which determines the confidence to be placed in them).

Firms are eager to obtain data since they want to improve production processes, develop innovative products and services, and more precisely target customers by adapting offers, advertisements and prices to them. Since each firm has the goal of moving ahead of its competitors, a race is on: who will make the best use of the data available? This has two major consequences on the demand for data. First of all, firms tend to pay a much higher price for data to which they have exclusive access than for data that they have to share with competitors. Secondly, competing firms within an industry might overinvest in acquiring and processing data, the effect being that they will eventually make lower profits. As in the prisoner's dilemma, competing firms would find it advantageous to collectively restrain their use of data, but no single player has an individual interest in doing so.

The supply of data

The data of value to firms come from three sources. First of all, many databases are freely accessible. The public sector produces most of these “open data” (*e.g.*, statistical, scientific or cartographical data). Universities and nongovernmental organizations have also opened access to their data; and even firms might find it worthwhile to do so (as in the case of Uber mentioned in the introduction). Secondly, firms also produce enormous amounts of data on their activities and the products they sell.⁶ Finally, you and I are, for sure, the biggest suppliers of the data of interest to firms owing to the previously mentioned datafication: we produce data either directly through our activities (the photos or “likes” posted on the social media, the websites visited, the e-mail sent, etc.) or indirectly through the devices we use (*e.g.*, connected wristwatches or smartphones with the geolocation feature turned). These data are valuable to firms insofar as they tell about our tastes, consumption patterns, social interactions, etc.

Is it possible, however, to talk about a “supply of data”? To pursue, we must examine this question. For a supply to exist, the access to data has to be controlled so that the producer is able

⁶ An autonomous car is estimated to generate up to 100 gigabytes of data per second, the equivalent of more than five million pages of text.

to set the conditions for an eventual transaction. In the case of the first two data sources, the producers (public organizations and firms) are quite capable of setting the conditions for access to their data. Access is deliberately public for open data; but for corporate data, it is usually closed or controlled under contractual arrangements, as we shall see.

But what about the data produced by individuals? Can we control access to our own data? In theory, yes.⁷ The websites we visit or the connected devices we use ask us to agree with their users' conditions. Even when there is the possibility of "opting out" and thus refusing to have our data collected, we seldom select this option. Why? First of all, we deem it costly (in time and effort) to read the terms and conditions or to apply the methods for restricting the collection of our data. It would take 76 days to read all the terms and conditions of use that an average American accepts to sign in a year (GRALLET *et al.* 2018). A second point, which partly justifies the first: we accept to obtain, in exchange for our data, services that are cheaper (often for-free), better adapted to our needs (via targeted advertisements) and potentially of better quality.⁸ This means that we assign a "virtual price" to our data and, thereby, our privacy.

This virtual price might turn into a real price, as when firms (for example, Internet access-providers) offer services to consumers for a price so that data not be collected on them or so that they not receive advertisements. By choosing this sort of offer, consumers show they are willing to pay in order to protect their privacy. This is an "opt out" method, since the consumer is the party who pays to close access to personal data. What happens if, on the contrary, it is up to the firm to pay consumers to open access to their data (an "opt in" method)? We tend to think that nothing would change at the same price (to be paid or received in payment) and for a given variation (up- or downwards) of the degree of protection provided. For a given level of data protection however, consumers generally demand in exchange for lesser privacy a monetary payment that is higher than what they are willing to pay to protect their privacy (ACQUISTI *et al.* 2013, SCHOLZ 2014). Consumers seem, therefore, to assign more value to their data when their consent is necessary for using them ("opt in") than for not using them ("opt out").

Relating supply and demand

Williamson (1991) has pointed out three ways of organizing economic organization: the hierarchy organizes transactions in an integrated firm; the market uses prices to coordinate supply and demand; and, between these two, a hybrid form that relies on contracts. Currently, data transactions are mainly organized in the first or third ways. In the first case, firms directly collect or produce the data they need.

The collection of personal data is based on "opt-out" contracts. Bluntly put, firms use the data as long as consumers do not keep them from doing so. As we have shown, vertical integration is justified when the data helps obtain a competitive edge, since firms have no interest in sharing their collected data and even less so the extracted information and knowledge. Legal restrictions on the sharing of personal data reinforce this tendency toward vertical integration.

Nonetheless, firms might find it in their interest to share their data in order to better coordinate activities, as in the case of the firms working together to develop self-driving vehicles. Transactions are then based on multilateral long-term contracts.

Another form of hybrid governance is to have recourse to middlemen specialized in collecting and processing data. These so-called "data brokers" propose customized services that firms appreciate when they cannot themselves collect data. Because of the aforementioned economies of scale and of scope, a few firms (mostly American) are dominant in the data brokerage business. They collect diverse data on millions of consumers worldwide: Acxiom (marketing), Equifax (insurance), Experian (loans), Corelogic (real estate) and Datalogix (finance).

⁷ To restrict access to our data, we could, for example, delete cookies from our browsers or use proxy servers.

⁸ For example, a connected meter adds no value unless it can accurately measure our consumption of water or electricity.

A “data market” as such does not yet exist however. True, there are a few platforms or data exchanges; but they are limited to a specific businesses and restrict considerably the transactions that can be made. The following paradox accounts for this. Since data are strategic, the proclivity to pay for nonexclusive data is usually low or inexistent. Given that data are nonrivalrous (their consumption by one party does not reduce the possibilities of another party consuming them), it is hard to guarantee that they are exclusive, even more so in a decentralized data exchange. In addition, it difficult to solidly establish the veracity of the data or their value, given that they are unique (no point of comparison) or are complementary (when several databases have to be combined to extract relevant information) (KOUTROUMPIS *et al.* 2017).

Conclusion

In summary, an ever growing quantity of data is being produced, collected and used; but only a limited fraction of it is exchanged. Three explanations have been provided: the strategic nature of data for firms, the difficulty of organizing decentralized data markets, and the lack of control by individuals over the data they themselves produce. With regard to this last factor, major changes are expected in the near future. A new EU regulation, the GDPR, is coming into force that requires firms to grant more control to individuals over their personal data.⁹ Firms will have to obtain positive, express consent from individuals in order to use their data and will also have to ensure the portability of these data — to see to it that consumers can take their data with them when they switch places of business. As Peitz and Schweitzer (2017) have explained, portability keeps the data from being locked down and facilitates competition for access to them (for want of a genuine secondary data market).

Another source of change is the emergence of intermediaries who propose solutions to consumers for actively managing their personal data and eventually monetizing them.¹⁰ The Facebook/Cambridge Analytica scandal has set off such strong reactions among cybernauts and public authorities that we can imagine that a new era is dawning: transactions involving personal data will be more regulated, more transparent and more respectful of individuals.

References

- ACQUISTI A., JOHN L.K. & LOEWENSTEIN G. (2013) “What is privacy worth?”, *Journal of Legal Studies*, 42, pp. 249-274.
- GRALLET G., PONCET G. & PONS H. (2018) “Données personnelles: comment reprendre le contrôle?”, *Le Point*, 25 January.
- KOUTROUMPIS P., LEIPONEN A. & THOMAS L. (2017) “The (unfulfilled) potential of data marketplaces”, *ETLA Working Papers*, 53. <http://pub.etla.fi/ETLA-Working-Papers-53.pdf>.
- PEITZ M, & SCHWEITZER H. (2017) “Datenmärkte in der digitalisierten Wirtschaft: Funktionsdefizite und Regelungsbedarf”, Discussion Paper No. 17-043 (Mannheim: ZEW).
- SCHOLZ E.M. (2014) “Putting a price on privacy: An introduction to the economics of privacy”, IPdigit.eu (1 June), <http://www.ipdigit.eu/2014/06/putting-a-price-on-privacy-an-introduction-to-the-economics-of-privacy/>.
- THIERAUF R.J. (1999) *Knowledge Management Systems for Business* (West Port, CT: Quorum Books).
- WILLIAMSON O.E. (1991) “Comparative economic organization: The analysis of discrete structural alternatives”, *Administrative Science Quarterly*, 36, pp. 269-296.

⁹ General Data Protection Regulation: “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data”. Available via: <http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1478961410763&uri=CELEX:32016R0679>.

¹⁰ The best known among these PIMS (personal information management systems) are Datacoup, Digi.me and Meeco.