

# Will Big Data Deliver its Promised Productivity Growth?

Lionel Artige\*

HEC - University of Liège

September 20, 2015

## Abstract

There are high economic expectations regarding the emergence of Big Data: a promised golden age for both consumers and firms. Digital technology allows firms to provide digital services in exchange of personal data, from which they can fine-tune their supply to better match market demand. By extending Arrow (1962)'s analysis of the market for information to the phenomenon of Big Data, we propose a theoretical assessment of its potential effects on productivity growth. Our study highlights that Big Data modify the market for information by introducing a new informational product, based on personal data, which is not a legally disseminable public good. This characteristic is crucial as it implies that private production of personal-data based information can be profitable. This profitability gives digital firms the incentive to produce information likely to improve firm-level productivity and consumer welfare. Finally, we conclude that the productivity effect of Big Data at the macroeconomic level is conditional on the returns of both the production and diffusion of this personal-data based information.

Keywords: Big Data, competition, competition policy, economic growth, personal information, productivity.

JEL Classification: D8, L1, O4.

---

\*HEC - Université de Liège, Economics department, Boulevard du Rectorat 7, B. 31, 4000 Liège, Belgium. E-mail: lionel.artige@ulg.ac.be. This research was funded through the ARC grant for Concerted Research Actions, financed by the French-speaking Community of Belgium. I would like to thank Laurent Cavenaile for our fruitful discussions and his comments on an earlier version of this paper. I also wish to thank Axel Gautier, Nicolas Petit, Mirjana Pejic Bach and participants of the ENTRENOVA 2015 conference for their valuable comments. Responsibility for all errors is our own.

# 1 Introduction

The volume of data created every year in the world started to increase sharply with the arrival of the personal computer but has stepped up considerably the pace with the Internet, the sensor and the mobile phone (Lyman & Varian 2003). This proliferation of data has been made possible by three technological innovations: processing, storage, and information and communication technologies. These innovations have changed our professional and free time activities in a convergent way. Most of our acts now proceed through screens and hard disks in white- and blue-collar jobs as well as in our recreational moments. These new work and playful tools are also new media through which all our acts are inexorably "dataficated". Digitalization necessarily implies the production of data in unprecedented quantity, systematically stored in the new magnetic and optical media. This massive creation of digital footprints is not only stored at low cost but also cheaply transmitted electronically through communication technology (telephone, radio and TV, and the Internet). The "datafication" of almost everything, the rapid electronic transmission and the interconnection of all these data characterize this technological phenomenon called "Big Data".<sup>1</sup>

There are high expectations of Big Data in science<sup>2</sup>, public administration and, particularly, in business<sup>3</sup>. Some consider data as a new raw material or even a new class of assets, i.e. goods capable of generating and storing economic value (The Economist (2010) and Schwab et al. (2011)). Many expect lower prices for consumers thank to ubiquitous competition, better-informed decisions for enterprises thank to data analysis, and more income growth and jobs at the macroeconomic level thank to higher productive efficiency.

Are these high expectations misplaced? Will Big Data deliver its promise of productivity growth? This paper attempts to answer these questions by extending the theoretical analysis of the market for information by Arrow (1962) to an important segment of the Big Data phenomenon: the production of personal digital footprints by digital users and their exploitation by providers of digital services. What do Big Personal Data change to the market for information? Our theoretical exploration concludes that Big Personal Data modify this market by allowing digital firms to make profits by extracting valuable information from personal data that consumers accept to transfer in exchange of the use of a free or paid digital service. Data protection laws<sup>4</sup> prohibit passing on personal data to the public arena. This legal bulwark against privacy violations not only protects the users of the Internet and other digital networks but also safeguards the digital firms' profits. Privacy indeed distinguishes personal information from information in general.

---

<sup>1</sup>See Chen et al. (2014) for a survey and Boyd & Crawford (2012) for a critical analysis of this phenomenon.

<sup>2</sup>Scientific knowledge is expected to benefit a lot from massive data production and processing in natural and engineering sciences (see, for instance, Feigelson & Babu (2012) and Marx (2013)) as well as in social sciences like economics (Einav & Levin (2013), Einav & Levin (2014), Taylor et al. (2014)).

<sup>3</sup>See, for instance, Manyika et al. (2011) and Lohr (2012).

<sup>4</sup>For an overview of data protection laws in the world see the website: [www.dlapiperdataprotection.com](http://www.dlapiperdataprotection.com).

The protection against the public dissemination of personal information is thus the guarantee for digital firms to appropriate the economic value they create from our personal data. If personal information were not protected, it could diffuse easily and costlessly as any other type of information and producers of information would not make any profit. Big Personal Data are therefore a game-changing technological innovation in the market for information. Digital technology allows to acquire countless personal data that each individual would communicate to only, if at all, a few trustworthy people. By making possible to capture profit from the production of personal information, Big Personal Data provide digital firms with big incentives to produce new market information based on personal data. The quality of market information should increase and so does corporate efficiency thank to better-informed management. However, most firms will have to pay for such personal-data based information or live without it. At the macroeconomic level, data-based management of firms should lift up aggregate productivity provided that the production of personal-data based market information is not a monopolistic rent of a handful of digital giants.

The rest of the paper is organized as follows. Section 2 provides definitions of data, information and Big Data. Section 3 explains the characteristics of the market for information. Section 4 presents what the emergence of Big Data change to the market for information. Section 5 deals with the potential macroeconomic effects of Big Data. Section 6 concludes.

## 2 Data, Information and Big Data

This section aims at providing definitions of data, information and Big Data. Definitions vary in the literature. In this paper, we will use the following definitions:

**What are data?** Data can be defined as distinguishable variables with no meaning *per se*. Floridi (2010) talks about data as “lacks of uniformity” in the real world, between two signals or two symbols. He takes the example of a red light against a dark background. This is a datum in the “real world” breaking uniformity with the dark background without a self-imposing intelligibility. Silence amidst the noise is also a datum as lack of uniformity.

Data can be stored as continuous variables (analogue data) or as discrete variables (digital data). In fact, digital data are a particular case among discrete variables: they are binary variables. As is well-known, digital data are aggregates of bits (binary digits) of 0 and 1. This means that all digital data use the same lexicon composed of only two signs (bits represented by two figures 0 and 1) but differ in their number of signs and in the order of these signs (syntax). This two-sign lexicon has made possible to store in the same way a sound, an image or a text. The universality of the digital technology has allowed to digitalize the storage of many data, i.e. the transformation of analogue data into digital data.

**What is information?** Anyone with no computer knowledge looking at a computer programme may find it meaningless or unintelligible. Yet data in the computer programme are assembled by its designer to be meaningful. Data become informative when data in our environment produce a meaningful signal or when we associate a semantic content to these data (Floridi 2010). In other words, information can be defined as data + meaning. Then, knowledge emerges from the accumulation of information. In the terminology of Floridi (2010), knowledge builds up from factual semantic information, which implies that the semantic content associated to data has proved to be true.

**What are Big Data?** There are countless definitions of Big Data and most of them stress the size and the exponential growth of data we have been observing in the recent past. Floridi (2012) points out the tautology in defining Big Data by their size. Nevertheless, the high volume of production is indisputably a characteristic of Big Data. Two other characteristics are generally put forward (Laney 2001): high velocity of data production but also of data expiration; and variety of data (structured, unstructured, text and multimedia data). This 3V's definition of Big Data is not operational for our theoretical analysis. In this paper, we will define Big Data as mainly the production of personal data and information, i.e. Big Personal Data. This characteristics distinguishes data and information before and after the digital technology. Digital technology provides a technical way to offer services at a distance, which requires the transfer and storage of personal data to connect people and connect people with goods and services. By using these digital services, we leave additional digital footprints of our personality. Anytime we use a digital media, we tacitly accept to transfer personal data on a server, which raises major heatedly debated privacy concerns (Solove 2004). The massive production of personal data resulting from the use of digital services characterizes the singularity and the novelty of the era of Big Data.

## 3 The Economic Value of Information

### 3.1 Conditions for a Social Welfare Optimum in a Market Economy

Let us start with an example to highlight the determinants and the pitfalls of the market for information. Imagine Sarah, an ice-cream seller, owning her ice-cream parlor in a holiday resort where the weather is very changeable. Since the weather is uncertain, she would like to hire less staff on rainy days and hire more on sunny days to maximize her profit. Suppose that, due to labor regulations, she has to make hiring decisions at least two days ahead of time and in a context where there is no scientific meteorological knowledge to make weather forecasts fairly accurately. The absence of weather forecasting information creates uncertainty for her, which will surely result in erroneous hiring decisions with negative consequences on her profit. Knowing weather forecasts would solve her planning

problem, maximize her profit and reduce the ice-cream price in a competitive ice-cream market. The producer and the consumer surpluses would be higher. The weather forecast is thus valuable information but it does not exist. Suppose now that a businessman spends a lot of money investing in meteorological research and development to produce seven-day weather forecasts with a good level of accuracy. Weather is still uncertain but now predictable thanks to his technological innovation. Assume also that the businessman is a monopolist and, hence, a price maker. There is a potential market demand for weather forecasting information since all the ice-cream sellers, at least, would be interested in buying it. The question is what the market price for weather forecasts will be. If the businessman succeeds in selling his weather forecasts to Sarah, she will disclose them to the extra waiters simply by informing them whether they are hired or not the forthcoming days. The relatives and friends of these waiters will consequently be informed about their schedule the next days thus diffusing the forecasting information at no cost. Sarah's nearby competitors could also have a chance to get informed without paying a penny. The weather forecasting information would keep on diffusing informally at no cost. The following incentive mechanism is in order: Sarah, the only person bearing the cost of the weather forecasts, will probably soon stop purchasing them driving their price down to zero; the businessman will stop spending on research because there is no market return to cover the cost of his inventing activity. Paradoxically, weather information is valuable for all ice-cream sellers and yet this potential high market demand does not turn into many buyers and a high price for the monopolist. How to explain this nonoptimal allocation of resources to information?

General equilibrium theory - the framework of neoclassical economics - provides the conditions under which a society can achieve a social welfare optimum when supply, demand and price levels are determined without centralized decision-making (Arrow & Hahn (1971) and Mas-Colell et al. (1995)). Among these conditions, perfect competition is the most well-known. There are three generally observed reasons why perfect competition fails to achieve a socially optimal allocation of resources: indivisibilities, inappropriability and uncertainty. Indivisibilities occur when it is not profitable to produce goods under a minimum scale due to fixed costs. When this happens, free entry of new enterprises in such a market - a condition for perfect competition - fails and the resource allocation is nonoptimal. Inappropriability describes the situation in which profits from the private production of a good cannot be fully captured by the producer. Public goods are subject to this failure because several people can consume the good at the same time (non-rivalry) and because it is difficult to prevent those people from consuming the good without producing it themselves (non-excludability). An example of such a good is information, which can be costly to produce, used by its producer and some others simultaneously, and arduous or even impossible to avert its disclosure. There is thus no incentive for any private production of this good and the resource allocation under perfect competition is inevitably nonoptimal. Uncertainty refers to incomplete information. Information may be incomplete because existing information is known to some and unknown to others (asymmetric information) or because not yet existing information is unknown to all (future information). When uncertainty is introduced in the general equilibrium framework, the socially

optimal resource allocation gives way because individual preferences are assumed to have a concave form (i.e., people enjoy consuming more of a good but grow weary of it each time they consume an additional quantity). In an uncertain environment, the concavity assumption of the utility function means that individuals are risk-averse. The result is that they will invest less in risky activities than the social optimum requires. Although they could buy contracts to be insured against these risks, they could not do it for any type of risk because the insurance market is incomplete due to moral hazard.<sup>5</sup>

### **3.2 Information Market: Optimal Use and Nonoptimal Production of Information**

In a celebrated paper, the American economist Arrow (1962) explains why these three motives - indivisibilities, inappropriability and uncertainty - create a market for information whose production is socially nonoptimal. First, he argues that uncertainty creates an economic value for information. When information is incomplete because existing information is not known to everybody, the utility or profits of the individuals who do not possess such information may be negatively affected due to uninformed decisions. Knowing such information reduces uncertainty and increases these individuals' welfare by replacing uninformed by well-informed decisions. This economic value of information gives rise to a market for existing information, in which individuals who ignore a piece of information are willing to acquire it from individuals who know it. Optimality of allocation requires that the price of this piece of information equals its marginal cost, i.e. its diffusion cost since the product on sale is existing information. Nowadays, the diffusion technology is universal, fast and cheap thank to our modern information and communication technology. Therefore, the marginal cost should be close to zero. Can marginal cost pricing be applied to the market for existing information? It cannot because of indivisibilities and inappropriability, two reasons at the origin of nonoptimal allocation of resources. This is the second point of Arrow. Information is an indivisible commodity. Either one knows it or does not. Moreover, as already mentioned, information is a non-rival commodity which can be diffused easily and at low cost. The owner of the piece of information has therefore no incentive to sell it and may prefer to keep it for herself. Nevertheless, information is also hardly excludable, which implies that any information, as soon as it is used, has a high probability of leaking and becoming public. This is good news for the optimality of use of information. Although indivisibilities and inappropriability are obstacles to optimal allocation of resources through marginal cost pricing, inappropriability leads to the social optimum all the same thank to the non-excludability of information and the cheap cost of its diffusion technology. This reduces the incompleteness of information for all and increases social welfare thank to more well-informed decisions.

---

<sup>5</sup>If an entrepreneur can sign an insurance contract, which covers the risk of the commercial failure of his new product completely, he will not have the incentive to provide the maximum effort to make it a commercial success. Due to this presence of moral hazard, no insurance company will accept to fully insure this entrepreneur.

However, what is a virtue for the use of information is a vice for the production of information. In his third point, Arrow explains that inappropriability, which pushes prices down to zero, now is a serious brake to the production of knowledge, i.e. the production of information or new ideas. On the market for existing information, the cost is limited to the cost of diffusion. On the market for the production of information, the cost includes not only the diffusion cost but also the cost of research to produce new information. Moreover, this production is an uncertain process. There is no guarantee that research will yield useful and profitable knowledge. If the producer of information has no guarantee to retain the benefits of its costly and successful research, he will never allocate resources to research in the first place. Arrow concludes that indivisibilities, inappropriability and uncertainty lead to a suboptimal amount of resources allocated to invention activity.

In a market economy, optimal production of information and optimal use of information are incompatible. On the one hand, optimal use of information requires a price close to zero, which would result in zero production of information. On the other hand, optimal production of information requires a return high enough to cover the research and development cost, which is best reached by not diffusing the produced information leading to a nonoptimal use of information. In order to solve this dilemma, the institutional solution (centralized decision making) that has been found grants legal protection (patents, copyrights, ...) to the inventors, providing partial excludability of information and royalties to the inventors. This solution aims at combining the creation of rewards for the production of information and its wide diffusion for an optimal use of it. The system is not perfect but allows to get out from the incentive trap of the market for information.

## **4 What do Big Data Change to the Market for Information?**

### **4.1 Less Asymmetric Information**

Recall Sarah, the owner of an ice-cream parlour in a holiday resort. Imagine her going shopping before the Internet to buy clothes. She enters many shops, talks with the shopkeepers, tries on clothes, compares the quality, the style and the prices of the items, ends up making a choice and pays for her purchases. Then, she meets her friends for a drink and shares her feelings about the clothes she saw and bought and the shops she visited. It took time for Sarah to visit the clothing shops, even though they might be located near to one another, find clothes she would like to wear and compare prices. Most probably, she missed items she would have liked to buy and cheaper ones she already bought in more expensive shops. Nevertheless, she made her best to find the clothes she wants to wear at the lowest prices conditional on information she had at the time of making her purchases. On the supply side of this market, the shopkeepers made their best to infer information about Sarah in a very short span of time to meet her demand and conclude the sales. But they do not have tangible data to make up that information.

They do not know her precise age, her family status, her job, her friends and relatives and her true preferences. Asking by the shopkeepers for these data would have been a violation of her privacy. This lack of information, or more precisely, the presence of asymmetric information on both sides of the market probably resulted in an imperfect match between the supply of the shops and the demand of Sarah. These data exist but are not symmetrically shared by both sides of the market.

A decade later, Sarah wants to buy new clothes but can now go shopping online thank to the arrival of the Internet. She enters in a search engine the product names, the brand names and the store names she is interested in. When she has selected the items she would like to buy, she can use a price comparison website to find the store offering the lowest price and look for customer reviews on the items and the stores. If Sarah wants to make the purchase, she has to register as a new customer to the virtual store she has selected and enter some personal data about her such as her name, contact details and credit card information. Then, she can write her own customer review and share her experience and feelings in social media. All Sarah's clicks while browsing the Internet have left digital footprints on the servers of the search engine, the shopping sites she visited, the price comparison website, the shopping site where she made her purchase and her comments, and the social media. Whereas the shopkeepers had to infer information about Sarah's preferences and budget constraint from her appearance and a couple of dialogues in the physical store, computers can exploit digital data (data analysis) left by Sarah while browsing the Internet to infer the same type of information. The more data Sarah leaves on the Internet, the more of them can be used to refine and check the accuracy of information about her tastes and purchasing power constructed by the customer profiling algorithms.

This example highlights what change digital technology brings about in the market for existing information: the acquisition and the storage of personal data. Personal information is protected by privacy and data protection laws in many countries in the world and cannot be legally made public. Governments possess a part of our personal data called administrative data. Every individual must by law provide the administration with her personal data to fulfill her citizen's duties (tax payments for instance) and to enjoy civil rights as well as social benefits. These data, which used to be stored on printed paper, now are stored on hard disks. They remain confidential as in the past due to data protection laws. Individuals and firms cannot have a legal access to these data. However, digital technology has developed a legal access to our personal data. Each visit of a website and every information uploaded on social media generate digital data about ourselves that can be used by the private owners of these websites to learn more about our preferences, habits and resources. Every time we make a click while browsing the Internet, we voluntarily accept to transfer confidential personal data to the private entities we pay a virtual visit. These data are protected by data protection laws to some extent so that private entities cannot make them public.<sup>6</sup> In other words, by using the Internet, individuals freely give

---

<sup>6</sup>The security of the storage of personal data cannot be technically guaranteed. As shown by the disclosures on the US National Security Agency's spying in 2013, countless telephone and internet com-

private entities access to personal data to an unprecedented scale in history.

Big Personal Data, defined as the systematic digital storage of personal data, provide private firms with large quantities of personal data, which can be converted into information about individuals' tastes and budget constraints. Big Personal Data thus modify the market for existing information in two ways. First, they provide massive quantities of existing information on individuals which was not available before digital technology. Second, thank to the universality of the Internet use, Big Personal Data provide statistical conditions to build personal information from personal data with a level of accuracy unknown before. Moreover, it allows for correcting or updating this information on a continuous basis. In other words, digital technology provide technical tools to reveal existing information on individuals. If enterprises master these technical tools, they can collect massive personal data to reduce asymmetric information on their market, make better-informed decisions and match their customers' needs more precisely.

## 4.2 Personal Information Is Not a Legally Disseminable Public Good

The fundamental changes Big Data introduce in the market for information is on the production of information. First, digital technology allows the creation of two new informational products: personal information and personal-data based market information. By accumulating data transferred by digital users, digital firms can produce personal information for each of their customers. As already mentioned, the value of that information comes from the reduction in asymmetric information. If all digital firms are able to produce personal information on their customers, many less can aggregate it to produce statistically reliable information on market demand and supply and capture real-time changes in both sides of the market. Obviously, the larger the audience of digital firms, the more accurate the inferred market information, and the higher the value of this informational product. Therefore, the economic value of personal-data based market information will be dependent on the audience of users. As explained in the previous section, information is an indivisible good: it is accurate or it is not. Therefore, only market information produced by digital giants will be valuable. For this reason, Google and Facebook, for example, will never charge for the use of their web sites since their economic value relies on the size of their audience.

The second fundamental change - by far the most important one - concerns the profitability of the private production of these two new informational goods. Arrow (1962)

---

munications in the world have been intercepted by the US administration. Many more governments probably behave the same way in order to have access to personal data accumulated by private entities. In general, governments including the US administration store but do not diffuse the intercepted data in the public arena. It is nevertheless a real threat for companies and individuals as exemplified by the disclosure of many sensitive documents of Sony Pictures Entertainment in 2014. The US administration suspects the North Korean government as the mastermind behind the hacking of the US company's servers and the diffusion of confidential data in the public arena.

showed that information as a public good could not yield a sufficient return to cover the investment cost of its producer. But the two new informational products introduced by the Big Personal Data phenomenon are not standard public goods. The first one - personal information - is a public good but cannot be disseminated, in principle, because the diffusion of personal data is prohibited by law.<sup>7</sup> The private producers of personal information can assist enterprises to target consumers and better match their demand. The second one - personal-data based market information - is a public good but can be very demand-specific. This informational product is an aggregate of personal information on a specific market. Its accuracy depends on the size of the audience of the producer of personal information. Digital giants have a definite comparative advantage in this market. Although the dissemination of this informational product is legally possible, their private producers can tailor it to the specific needs of each firm as well as sell market information to some firms first. Thus, thank to protected personal data, private producers of personal-data based information can confidently capture the returns on their investment. As a result, the market for information based on personal data escapes the incentive trap which characterizes the market for general information.

## 5 Will Big Data Deliver its Promised Productivity Growth?

As mentioned earlier, Arrow (1962) concluded that the market for information as a public good would lead to a suboptimal investment in the production of knowledge (i.e. factual semantic information Floridi (2010)), hence yielding lower productivity growth than its optimal trend. In order to correct this market failure in the information market, Arrow advocated for a legal framework protecting the economic return of information producers by granting them intellectual property rights (patents and copyrights) while fostering unlimited fair competition in the market for tangible goods and services. The competition policy framework in the western world has relied on the recognition of these two dimensions of many products: the idea and the physical product. Competition is the rule for the production of the physical good or service by any enterprise abiding by the patents and copyrights while the producer of the idea benefits from an exception to this rule by receiving a guaranteed return to his invention. This regulation system is not perfect but has found a subtle balance to provide the right incentives to inventors (producers of knowledge) and maximize social welfare through competitive prices on the market of physical products. How does digital technology modify this balance?

On the consumer side, information and communication technology transforms any consumer into an online ubiquitous consumer. Prior to the Internet, price comparison was a costly and time-consuming activity. Nowadays, any web shopper can easily compare prices of a good he wants to buy before the purchase. This transparency of the market

---

<sup>7</sup>Obviously, in social media websites, personal information becomes public information if the users configure their profile accounts as public.

supply has inevitably increased competition and, hence, lowered prices. In this respect, social welfare has increased without destabilizing the competition regulatory balance.

On the producer side, digital technology has two main effects. First, it transforms tangible products into intangible ones. The most famous example is the music industry. All of a sudden, the physical dimension of a musical record vanished, transforming a rival and excludable good into a non-rival and non-excludable one. Artists and music corporations lost the actual property of the returns of their productions. In fact, digitalization of tangible goods and services in an increasing number of industries threatens the existence of businesses and jobs. Musicians and record labels have not disappeared but their revenues have decreased substantially on average despite a record demand for now very affordable music listening.

However, the producers have taken their revenge on the consumers by collecting and storing personal data of the frenetic web consumers. This is the second effect of digital technology on the production side: the emergence of Big Data. As explained in the previous section, Big Data do not eliminate market uncertainty for firms but reduces some of it. Some of the uninformed decisions that enterprises used to make amid the absence of information or based on biased scattered information should now be made based on market information statistically built from Big Data. The management of enterprises can therefore rely more on statistical personal-data based information rather than on uncertain anecdotal information or on the subjectivity of their managers. On average, productivity of firms should thus increase.

The competition regulatory balance is shattered by the two effects of digitalization on the supply side. By eliminating the physical dimension of some goods and services, digital technology threatens the profitability of private production of the digitalized goods. In the past, competition authorities had to regulate producers to the benefit of consumers. Nowadays, consumers should be regulated to make sure that producers get even a slight return of their production. The second effect of digital technology - the emergence of Big Personal Data - re-balances the information market in favor of the producers, which is good news for the industries experiencing the digital transition of their production. However, Big Personal Data do not solve the issue, highlighted by Arrow, of the incompatibility between the optimal production of information and the optimal use of information. Market power just changed sides from the users to the producers. Therefore, the macroeconomic effect of Big Data is uncertain. The competition policy framework will have to find a new subtle balance between the incentives for personal-based information production and its widest possible diffusion before privacy engineering technology eliminates the access to our personal data for all digital companies, i.e. the source of their market power.<sup>8</sup>

---

<sup>8</sup>There is a competition policy debate about Big Data. See, for instance, Stucke & Grunes (2015).

## 6 Conclusion

The objective of this paper is to propose an assessment of the potential effects of Big Data on aggregate productivity. The method of analysis is inspired by Arrow (1962)'s theoretical study of the market for information and its macroeconomic consequences. We extend his analysis by investigating the changes digital technology brings about in the market for information and its implications for the economy as a whole. Our main conclusions are threefold. First, we emphasize that Big Data provide digital firms with personal data as raw materials to produce personal information likely to improve firms' management decisions and better match their customers' needs. Second, data protection laws safeguard both the digital users' privacy and the digital firms' profit. If personal data were not protected, digital users would not trust digital firms and, consequently, would stop using their digital services to avoid transferring their personal data. It is in the interest of digital firms not to diffuse personal data because it is a condition for getting users, producing personal information and selling personal-data based information at a profitable price. This legal protection of privacy implies that personal information is not a legally disseminable public good. This is the reason why private production of information based on personal data can be profitable unlike general information which can be diffused easily and costlessly. Finally, we argue that Big Data should benefit all firms by reducing asymmetric information. However, the effect on the economy-wide productivity growth will depend on the level of competition prevailing in the production of personal-data based information. If competition is too high, prices will be low and digital firms will have no incentive to produce information. If competition is too weak, prices will be high and many firms will not be able to afford acquiring such information. Big Data's promised productivity growth is therefore conditional on both the production level and quality as well as the wide diffusion of this personal-data based information. If privacy engineering technology happened to succeed in ringfencing the access to our personal data for any digital service provider, the production and the diffusion of market information as well as production efficiency would likely be impacted negatively.

## References

- Arrow, K. (1962), Economic welfare and the allocation of resources for invention, *in* ‘The Rate and Direction of Inventive Activity: Economic and Social Factors’, National Bureau of Economic Research, pp. 609–626.
- Arrow, K. & Hahn, F. (1971), *General competitive analysis*, Holden-Day.
- Boyd, D. & Crawford, K. (2012), ‘Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon’, *Information, communication & society* **15**(5), 662–679.
- Chen, M., Mao, S. & Liu, Y. (2014), ‘Big data: A survey’, *Mobile Networks and Applications* **19**(2), 171–209.
- Einav, L. & Levin, J. (2014), ‘Economics in the age of big data’, *Science* **346**(6210), 715–721.
- Einav, L. & Levin, J. D. (2013), ‘The data revolution and economic analysis’, *NBER Working Paper* (19035).
- Feigelson, E. D. & Babu, G. J. (2012), ‘Big data in astronomy’, *Significance* **9**(4), 22–25.
- Floridi, L. (2010), *Information. A Very Short Introduction.*, Oxford University Press.
- Floridi, L. (2012), ‘Big data and their epistemological challenge’, *Philosophy & Technology* **25**(4), 435–437.
- Laney, D. (2001), 3d data management: Controlling data volume, velocity, and variety., Technical report, META Group.
- Lohr, S. (2012), ‘The age of big data’, *New York Times* .
- Lyman, P. & Varian, H. (2003), How much information? 2003, Technical report, Berkeley School of Information.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A. H. (2011), Big data: The next frontier for innovation, competition, and productivity, Technical report, McKinsey Global Institute.
- Marx, V. (2013), ‘Biology: The big challenges of big data’, *Nature* **498**(7453), 255–260.
- Mas-Colell, A., Whinston, M. D. & Green, J. R. (1995), *Microeconomic theory*, Oxford University Press.
- Schwab, K., Marcus, A., Oyola, J. R. & Hoffman, W. (2011), Personal data: The emergence of a new asset class, Technical report, World Economic Forum.

- Solove, D. J. (2004), *The digital person: Technology and privacy in the information age*, New York University Press.
- Stucke, M. E. & Grunes, A. (2015), 'Debunking the myths over big data and antitrust', *CPI Antitrust Chronicle* pp. 1–10.
- Taylor, L., Schroeder, R. & Meyer, E. (2014), 'Emerging practices and perspectives on big data analysis in economics: Bigger and better or more of the same?', *Big Data & Society* **1**(2), 1–10.
- The Economist (2010), 'Data, data everywhere', *The Economist* .