



Data sharing practice in Big Data ecosystems

Laia Pujol Priego

David Osimo

Jonathan Wareham



ESADE Working Papers Series

Available from ESADE Knowledge

Web: www.esadeknowledge.com

© ESADE

Avda. Pedralbes, 60-62

E-08034 Barcelona

Tel.: +34 93 280 61 62

ISSN 2014-8135

Depósito Legal: B-4761-1992

DATA SHARING PRACTICE IN BIG DATA ECOSYSTEMS

Laia Pujol (Ramon Llull University, ESADE)
laia.pujol@esade.edu

David Osimo (The Lisbon Council)
david.osimo@lisboncouncil.net

Jonathan Wareham (Ramon Llull University, ESADE)
jonathan.wareham@esade.edu

ABSTRACT

Big data has gained popularity in recent years both in industry and academia given actual and potential impact on contemporary business organizations. The combination of datasets, both internal and external to an organization, and the reuse of data for different purposes have been pointed to as the greatest value of big data (Davenport, 2013; Barton & Court, 2012). However, little is known (Thomas & Leiponen, 2016) about how firms, and particularly traditional industries, actually share data and which mechanisms they use to do so. While Open Innovation literature has already comprehensively captured the use of the various organizational modes through which external sources of *knowledge* are combined with internal developed knowledge (Chesbrough, 2003a, 2003b), they have not yet been applied to the understanding of *data* sharing practices in big data ecosystems and how industry characteristics affect the modes that companies use to share data in traditional industries. From a purposeful sampling of 102 real-life cases covering corporate early adopters of Big Data in eleven traditional sectors, this paper develops a conceptual framework of organizational modes implemented by traditional business in the big data ecosystem to extract business value from big data.

Keywords: Big data, Big data ecosystems, Data sharing, Traditional businesses, Open Innovation.

1. Introduction

Today, sensors in mobile phones, factories, cars, lorries, aeroplanes, shops and streets are gathering data at previously unheard-of levels of granularity in all industries, included those who used to be non-data intensive: The so-called *Internet of Things (IoT)* was estimated in 2015 to include 16.3 billion¹ devices or 2.2 devices per person connected to the Internet, transferring more than 1² zetabyte of data (Zikopoulos et al., 2012). This explosion of on-demand data is what has been recently dubbed *Big Data*, which describes high-volume, high-velocity and high-variety information assets that are useful for enhanced insight and decision making (Gandomi & Haider, 2015). *Volume* refers to the quantities of big data; *velocity* bears on the speed of data collection, processing and real-time analytics; and *variety* means the wide range of typologies of data gathered (McAfee et al., 2012; Hartmann et al., 2016). But these famous and widely used three *Vs* to describe big data—recently extended to five with the addition of *value* and *veracity* (Demchenko et al., 2013)—offer only a partial understanding of what it is.

Big data refers not only to (1) *data*, generally defined as qualitative or quantitative factual statements usually collected, processed, stored and transmitted over digital information infrastructures and/or elaborated with digital technologies (OECD, 2011, 2013); but also to (2) the *technology* and computing infrastructure required to rapidly aggregate, manipulate and process huge volumes and varieties of data (McAfee et al., 2012; Demchenko et al., 2013); and (3) the *analytical techniques and capabilities*, which refer not only to “agile” and advanced analytical techniques (e.g. time series analysis, machine learning, predictive modeling), but also to big data mindset, or “ideas about ways

¹ The short-scale [*échelle courte*] billion is used throughout (10^9) rather than the long-scale [*échelle longue*] billion (10^{12}). The short scale reflects traditional British and Continental usage.

² To give an inkling of this vast quantity, 1 Zetabyte = 10^{15} gigabytes or 10^{21} kilobytes [International System of Units - *Système international d'unités* — SI]

to tap data to unlock new forms of value” (Mayer-Schönberger & Cukier, 2013; Manyka et al., 2011; Wamba et al., 2015).

Whether data, technology or analytical expertise, firms often require working with other stakeholders to extract value from big data (McAfee et al., 2012; Davenport, 2013). *Value* is understood in this context as the term is used in the Information and Technology (IT) literature for IT investments to describe how economic benefits for the organization are captured through big data implementation (Parmar et al., 2014; Manyka et al., 2011; Devaraj & Kohli, 2003; Mithas, Tafti, Bardhan, & Goh, 2012). The value of big data is created along the *big data value chain*, which refers to all activities needed from data generation until data usage to extract value from big data (European Commission, 2013; Miller and Mork, 2013). Different types of organizations interact in the data value chain creating the so-called *big data ecosystem*, which extends the concept of value chain to that of a system that integrates any organization that takes part in the shared offering (Cattaneo et al., 2016; Curry, 2016; Autio & Thomas, 2017; Iansiti & Levien, 2004a, b).

The big data ecosystem integrates different players, which include traditional businesses, who are *data holders*, which refer to organizations that produce data and eventually provide their data to other third parties to reuse them (Thomas & Leiponen, 2016). Traditional businesses gather (and can eventually provide) their internal data from IT systems (e.g. ERP, CRM); exhaust data produced by unrelated business transactions, which refers to data that are shed as a byproduct of people’s actions and movements in the world (Mayer-Schönberger & Cukier, 2013); data from sensors or physical devices; data generated from crowdsourcing or web collaboration; and data from business partners, consumers and suppliers (Cattaneo et al., 2016).

Merging datasets from many sources, both internal and external to an organization, is what has been pointed to as the greatest value of big data (Davenport, 2013; Barton &

Court, 2012). Data's value has shifted from primary use (i.e. using data for the purpose for which they were collected) to potential future uses (Opresnik & Taisch, 2015; Zhu & Madnick, 2009). For instance, the Spanish telecommunications company Telefonica not only gathers data on the mobility of its mobile phone clients in order to manage its infrastructure and traffic, but also reuses these data to provide intelligence services to other companies, such as retailers. These kinds of transformative uses affect how businesses value the data they hold and their decisions about who is allowed access to their data.

Thus, the spotlight in big data is put on the way companies *share* data, which refers to how companies make their data available through appropriate information interfaces to any other third party, whether through a contract or standard to which all parties agree, or by making data freely open (Stefansson, 2002). Obviously firms in many industries have shared information in the past; well-known examples are insurance underwriters' laboratories, banking, energy and telecoms, where the exchange of information is a critical aspect for carrying out their businesses. For decades market research firms and similar specialized organizations have aggregated industry data, but the difference is that data are seen today as "a raw material entering the marketplace; an asset independent of what it had previously aimed to measure" (Mayer-Schonberger & Cukier, 2013).

Although the literature highlights the great value of data sharing between firms in the big data ecosystem (Davenport, 2013; Barton & Court, 2012), little is known about how firms, and particularly traditional industries, actually share data and which mechanisms they use to do so (Thomas & Leiponen, 2016).

Open Innovation (OI) literature has already comprehensively captured the use of the various organizational modes through which external sources of knowledge are combined with internal developed knowledge (Chesbrough, 2003a, 2003b). While OI literature has

developed modes for knowledge sharing (Chesbrough & Crouther, 2006; Gassmann & Enkel, 2004), such modes have not yet been applied to the case of *data* sharing and reuse in big data ecosystems. Since knowledge can be owned, protected and traded in ways that are not available to data (e.g. IPR, patents, copyright) (Thomas & Leiponen, 2016), the application of the organizational modes from OI literature to understanding data sharing practices in big data ecosystems might differ and augment OI concepts. On the other side, industry-level characteristics might affect the use of diverse organizational modes in various traditional sectors, as happens with selection of organizational modes for knowledge sharing across industries (Christensen et al., 2005; Chesbrough & Crowther, 2006).

Hence, the present study seeks to answer:

- *How do traditional businesses share data in the big data ecosystem?*
- *What are the mechanisms in big data ecosystems that traditional businesses use to share data?*
- *What are the main patterns across traditional industries?*

By addressing these questions, this paper seeks to provide a conceptual framework of organizational modes that traditional businesses implement in the big data ecosystem to share data. The paper contributes to the OI literature by applying the different organizational modes from OI to another setting (i.e. big data ecosystems), while providing insights on the industry characteristics that affect organizational modes. Regarding the managerial contributions, the paper aims to contribute to the recent policy discussion of data market policies, which seek to promote data sharing practices among businesses, with special focus on traditional industries, but lack an analysis of the actual practices and bottlenecks. For businesses, it provides a systematic understanding of the data sharing practices and business value, in the big data ecosystem, which is claimed to

be one of the most promising transversal technology trends in the coming years that will enable companies, if they engage thoughtfully, to derive significant economic benefits.

The paper is organized into six main parts: Section 2 covers the background. Section 3 sets out the approach used for answering the research questions. Section 4 describes the results from the data analysis. Section 5 discusses the research results. Finally, Section 6 provides conclusions and suggests future lines of research.

2. Background

2.1. *From the big data value chain to big data ecosystems*

Value chain is a term that has been used in the business management literature to define the sequential activities that organizations implement in order to deliver a valuable product or service to the market (Porter, 1985). As an analytical tool, the value chain concept has been applied to understanding value creation in big data. The European Commission refers to the big data value chain as the “center of the future knowledge economy, bringing the opportunities of the digital developments to the more traditional sectors” (European Commission, 2014).

Different taxonomies have been used to describe the sequential activities in the big data value chain (e.g. Cattaneo et al., 2016; Thomas & Leiponen, 2016; Filippov, 2014), which could be grouped into the following: (a) *Generation and collection*, or the process of data gathering; (b) *Storage, aggregation and organization*, which describe the activities related to facilitating the persistence and management of data over the life cycle in a scalable way to enable applications to have fast access to the data (Pennock, 2007); (c) *Analysis*, which includes the activities related to transforming the data in a decision-making and domain-specific usage; and (d) *Usage*, which the literature further divides into primary use, when data are used for the purpose for which they were gathered, and

secondary use or reuse, when data are exploited for purposes other than those for which they were initially collected.

Different types of organizations interact in the data value chain in order to generate value from data, forging the so-called *big data ecosystem* (Cattaneo et al., 2016; Curry, 2016). The ecosystem metaphor has been applied to describe big data environment where companies combine their individual offerings into an integrated solution, which allow firms to create value by joining efforts (Adner, 2006; Moore, 1996; Wareham et al., 2014). The term *ecosystem* was coined by Tansley (1953) to identify a basic ecological unit comprising both the environment and the organisms in it. It was later adopted using the biological analogy for business contexts to describe an “economic community supported by a foundation of interacting organizations and individuals” (Moore 1993, 1996). The organizations in the big data ecosystem (Figure 1) include (Thomas & Leiponen, 2016; Cattaneo et al., 2016, Hammel et al., 2012; Curry, 2016):

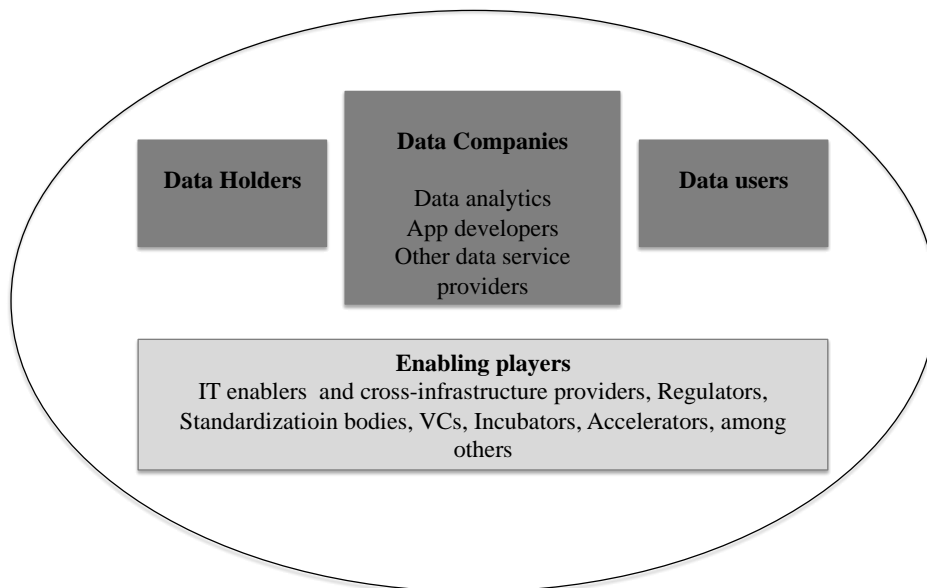
(a) *Data holders*, organizations that generate the data and eventually provide their data to third parties for reuse.

(b) *Data companies*, firms whose main activity is the production and provision of data-related products, services and technologies. This category includes *data analytics companies*, who provide an extensive variety of products such as analytics platforms, business intelligence, artificial intelligence, etc.; *application developers*, who design, build and sell applications that enable the commercialization of big data, including vertical solutions, mobile apps, cloud apps and big data apps; and other *data service providers* who are specialized in developing and selling services based on the reuse of data, whether cross-sector or specialized in specific vertical markets; these also include data aggregators, specific service providers specializing in the collection and aggregation of big data.

c) *Data users*, organizations from different industrial sectors that use big data technology and services for their own internal exploitation or to produce data-based products and services for their own customers.

d) *Enabling players*, which include, among others, Information and Communication Technology (ICT) enablers and cross-infrastructure providers; regulators for data privacy and legal aspects; standardization bodies that define de jure or de facto technology standards to promote widespread adoption of big data technology; investors, venture capitalists, incubators, accelerators, and governmental funds that facilitate access to risk capital.

Figure 1. Big data ecosystem (adapted from Cattaneo et al., 2016; Curry, 2014; Hammel et al., 2012)



Traditional businesses in the big data ecosystem are the *data holders*, the ones owning or generating the data. They can vertically integrate the complete value chain from generation to use of data, or rely on collaboration with other organizations and specialized data companies to cover the steps of the value chain. Over time, the value is shifted to those who hold the data, the data owners, who are the object of the present study (Mayer-Schonberger & Cukier, 2013).

2.2. *Traditional businesses in big data ecosystems*

According to the latest data published by the European Commission, the uptake of big data in traditional sectors is still limited: Only about 6% of companies are adopting big data solutions, and adoption varies greatly depending on the industry, with finance, wholesale and retail sector and utilities—industries that are IT and data intensive—on the side of the spectrum with highest adoption, and construction, transport and the public sector on the other side, confirming their slow pace in adopting big data (Cattaneo et al., 2016).

The differences between big data adoption across sectors are also reflected in the distribution of data workers across industries. *Data workers* refer to those who collect, store, manage and analyze data as their primary, or a significant part of, their job activity (European Commission, 2014). Four industries—manufacturing, wholesale and retail, professional services and Information and Communication Technologies—accounted for nearly 62% of data workers in 2016. Sectors with the lowest concentration of data workers were construction, transport and healthcare.

Businesses in traditional sectors can extract *business value* from big data in different ways (Manyika et al., 2011; Banerjee et al., 2011; Piccoli & Pigni, 2013; Parmar et al., 2014). These include: (a) *productivity and efficiency improvements*, by reducing or avoiding operational costs due to process automation or improved risk management (Anderson, Fornell, & Rust, 1997); (b) *segmentation*, by facilitating the personalization of offerings and services through enhanced knowledge of customers' needs and shaping customer behavior (Gopalkrishnan, Steier, Lewis, Guszczka, & Lucker, 2013; Ansari and Mela 2003; Mithas et al., 2012); (c) *improved product or service*, by changing an existing good or service through increasing the degree to which that product or service is successful in producing a desired result (Cattaneo et al., 2016); and (d) the *creation of a*

new product or service, by generating a new offering (product or service) in the market thanks to big data implementation (Krishnan & Ulrich, 2001; Gopalkrishnan, Steier, Lewis, Guszczka, & Lucker, 2013).

In order to extract business value from big data, traditional businesses are reaching out to other stakeholders in the big data ecosystem and are experimenting with different organizational modes to obtain data, technology or analytical skills to cover the overarching activities in the big data value chain.

2.3. What Open Innovation can reveal about traditional businesses practices in big data ecosystems

The big data ecosystem is constantly evolving, which is precisely a defining characteristic of innovation ecosystems that adapt and evolve over time (Basole, 2009; Weill and Woerner, 2015). In big data ecosystems, firms are part of a system characterized by distributed knowledge in which data, technology and analytical capabilities are exchanged in order to extract business value from big data. OI highlights the importance of external sources of knowledge combined with internal developed knowledge (Chesbrough, 2003a, 2003b) in a process where innovation is distributed across several actors in the innovation system (Tether, 2002; Acha and Cusmano, 2005). OI requires companies to engage in innovation ecosystems that include a variety of actors throughout various steps of the innovation process (West and Bogers 2014).

Scholars have systematically identified the various mechanisms by which firms combine external and internal knowledge to innovate in innovation ecosystems (Vanhaverbeke et al., 2014; Lichtenthaler, 2004; Lichtenthaler, 2005; Hughes & Wareham, 2010); and have specifically looked at how companies in traditional industries applied them, showing that firms in traditional industries engage in OI practices differently from companies in high-tech sectors (Chesbrough and Crowther, 2006). The

mechanisms to implement OI by combining external and internal knowledge take the form of diverse organizational modes such as: partnership, joint venture, acquisition, and a long list of other organizational forms ranging from single partner relationships to multiple or community based collaborations (Chesbrough & Bogers, 2014).

While trying to apply OI as theoretical lenses to understand data sharing practices in big data ecosystem, it is important to understand the differences between knowledge and data (e.g. Bhimani and Willcocks, 2014). Besides the widespread and equally contested hierarchical view, which advances from data (facts), to information (processed data), to knowledge (personalized information) (e.g. Tuomi, 1999, Alavi & Leidner, 2001; Dreske 1981), data are non-rivalrous and partially exclusive (Mayer-Schonberger & Cukier, 2013; Thomas & Leiponen, 2016; Pantelis & Aija, 2013). Data can be copied, data value do not diminish when they are used as they can be processed many times by several people, and they are hardly legally protected (Thomas & Leiponen, 2016). Data, like other goods, are traded depending on their quality, potential users' willingness to pay and price (Pantelis & Aija, 2013). The value of data needs to be considered in terms of all possible ways the data may be employed in the future, not only about their present usage; and when multiple datasets are combined together, the result is more than the sum of individual components (Mayer-Schonberger & Cukier, 2013). If we consider these attributes of data, then the importance of contractual agreements regarding data ownership, data usages and data value becomes clear (Manyka et al., 2011; Chui et al., 2013).

3. Method

3.1. Sample

Our research design has used an inductive approach in accordance with the exploratory nature of the study. We implemented a thematic analysis of qualitative data from publicly available sources from a purposeful sample of 102 real-life cases covering eleven sectors: Aerospace, Agriculture, Automotive, Chemical, Energy, Finance, Health, Machinery, Retail, Telecommunication, and Transport. Those eleven sectors were selected to deliberately seek variance in *data intensity*, meaning the level of capability and sophistication of big data use in a sector (Manyika et al., 2011). They were chosen in alignment with the Data Market Monitor³ commissioned by the European Commission in order to enable us to access recent quantitative data on big data per industrial sector, to be triangulated with the present study to shed light in some of the patterns identified.

Since big data is a fast-evolving trend, with growing but limited adoption among companies (Cattaneo et al., 2016), the selection of 102 real-life cases deliberately focuses on *early adopters*, companies in traditional sectors at the forefront of big data technology adoption in each of the traditional sectors. In other words, the cases are not selected to be representative of the population, but to have precedent-setting value: By looking at early adopters we are able to identify issues and behaviors that might later be extended to the overall population of companies adopting big data. Two of the dimensions are constant in all cases—private companies (large or SMEs), and early adopters of big data technology—while variation exists in two other dimensions, namely the specific sectors: Aerospace (N=6), Agriculture (N=11), Automotive (N=9), Chemical (N=9), Energy (N=8), Finance (N=9), Health (N=10), Machinery (N=11), Retail (N=10),

³ European Monitor for Data market: <http://www.datalandscape.eu/>

Telecommunication (N=10), Transport (N=9); and geography (the company's country of origin), which includes Europe, US, Latin America and Asia. In order to identify the companies, we employed existing general and sector-specific reports that typically include exemplary cases.

3.2. *Data collection and analysis methods*

In the preparatory phase of the study, as a first step before the selection of cases, six semi-structured interviews were carried out with sector experts and managers of data analytics companies in order to reach a better understanding of the main actors and mechanisms for data sharing practices in big data ecosystem, and to inform the subsequent sampling strategy.

In a second phase, data on the 102 selected cases were collected using secondary data from publicly available sources, the most important ones being company websites, consultancy reports and additional data from news sources. In total, 133 different sources were examined in order to analyze the 102 cases. Sources for each of the 102 cases were manually searched to identify statements about the use of big data in their business and were coded accordingly. The selection of sources per case was concluded when it was determined that no significant additional insights were captured from additional data points and theoretical saturation was reached.

Code concepts and categories emerged by triangulating prior literature and theory reviewed in §2 (piori coding) and the analysis of the archival data from all cases (open coding). Piori codes, detailed in §2, include:

- Activity in data value chain (Cattaneo et al., 2016; Curry, 2016), including (a) Generation and collection, (b) Storage, aggregation and organization (c) Analysis, (d) Usage.

- Business value (Piccoli & Pigni, 2013; Parmar et al., 2014), including (a) productivity and efficiency improvements (Anderson, Fornell, & Rust, 1997), (b) segmentation and shaping customer behavior (Gopalkrishnan, Steier, Lewis, Guszczka, & Lucker, 2013; Ansari & Mela 2003; Mithas et al., 2012), (c) improved product or service (Cattaneo et al., 2016), and (d) creation of a new product or service (Gopalkrishnan, Steier, Lewis, Guszczka, & Lucker, 2013).
- Organizational modes (Chesbrough and Crowther, 2006; Bianchi et al., 2011; Vanhaverbeke et al., 2014), which in the setting of big data ecosystem describe: (a) In-house, when there is no agreement with external organizations to perform any data-enabled process or service; (b) Acquisition, when the company acquires another company in order to gain ownership of their data and data-enabled services; (c) Outsourcing, which is further distinguished as (1) outsourcing data analytics, when a company outsources the big data analytics to match their data-gathering and (2) outsourcing data gathering, when a company outsources the collection of internal data; (d) Partnership, when data holders partner with specialized analytics companies to provide additional services to their customers; and (e) Joint venture, when a new business entity is created by two firms through investing equities and sharing members on the board of the new firm.

From open coding, salient themes not yet encompassed in the OI literature were identified, completing the codes regarding the organizational modes, which added: (f) Data reselling, when companies monetize data about their company through data platforms, which splits data revenue among firms; (g) Innovation space, when data is shared between selected organizations in a controlled and protected space provided by third-party organizations; (h) Community data, when data is shared in a peer-to-peer fashion, as a community where data is accessible only to those that provide their data too;

and (i) Open data, when data is shared openly, generally for purposes of marketing or communication. From the open coding process, the concepts related to organizational modes were clustered in the following categories: closed modes, where organizations vertically integrate data value chain and keep data ownership and usage inside the company boundaries (includes acquisition, in-house and outsourcing); shared modes, where organizations disclose their data to third parties to use it through contracts or standards to which parties agree (data reselling, innovation space, joint venture, partnership); and open modes, where data is shared openly (community data and open data). The complete code features are provided in Appendix 1.

All sources were coded by two independent coders (co-authors of the present paper) to ensure reliable coding. The results of the coding performed by the two coders were compared; differences were discussed and disagreements were resolved. There was 95% agreement, with the 5% disagreement concentrated in the categories of business value. The codes covering business value were refined after the meeting to avoid any ambiguity.

In a third phase, the study implements a cross-level analysis approach, where the results from the analysis of previous phase focusing on firm practices for sharing data are analyzed under an industrial perspective across sectors, responding to calls to provide cross-level analysis that combine firms and industry characteristics (Bogers et al., 2017). In the context of the big data ecosystem, the mutual dependence between the ecosystem and the industry in which it is implemented requested such analysis. To do so, and according to literature and theory reviewed in §2, we analyze how the mechanisms used by companies to share data (or not) are associated with *industry data intensity*, meaning the level of capability and sophistication of big data use in a sector (Manyika et al., 2011). We use as a proxy IDC data (Cattaneo et al., 2016) concerning *share of data workers in total employment*, and divide sectors in two categories: high data intensive (more than 5%

of the workforce are data workers), which includes finance and telecom; and low data intensive (below 5%), which includes retail, machinery, health, energy, and transport (some sectors in our analysis are not included due to lack of data availability in the IDC study). Obviously, the data intensity qualification is attributed to the sector and not to the individual company.

Finally, insights resulting from the analysis (phase 2) were cross-validated in a workshop held on the 21st of September, 2016, with 50 selected participants across the 102 real-life cases and sectorial experts, in the framework of a European Commission seminar, in order to present the interim results of the research and seek respondent validation from participants (Maxwell, 2012). The results from the workshop foster greater understanding of the implications of the mechanisms for exploiting data and insights on the challenges and bottlenecks in sharing and re-using data.

4. Results

a) Data sharing by traditional industries is uncommon

The frequency of organizational modes in the sample (Table 1) shows that the general pattern is one of closed data (89 of the 102 cases), where data of traditional companies is not shared with third parties outside the organization in the ecosystem. In 11 cases, the data were shared, and in only 2 cases were the data completely open.

Table 1. Frequency organisational modes

Category	Organisational mode	N. cases
Closed	Total closed	89 (87%)
	Acquisition	7
	In-house	17
	Outsourcing data analytics	38
	Outsourcing data gathering and analytics	27
Shared	Total shared	11 (11%)
	Data reselling	1
	Innovation space	1
	Joint Venture	3
	Partnership	6
Open	Total open	2 (2%)
	Open Data	1
	Community sharing	1
Total general		102

Looking at the specific organizational modes most commonly found in the sample, in decreasing order are: outsourcing data analytics (38), outsourcing both data gathering and analytics (27), in-house (17), acquisition (7), partnership (6), joint venture (3), open data (1), community sharing (1), innovation space (1), and data reselling (1).

An example of each mode is outlined in the table below:

Table 2 Examples per organisational mode

Organisational mode	Examples
Acquisition	A large agrochemical producer acquired a company owning large datasets and predictive models about weather and crop growth to open up the new “precision agriculture” market.
In-house	A telecom company developed in house a solution to offer value-added services to retailers about customer mobility based on the telecom operator data from mobile phones.
Outsourcing data analytics	A large chemical manufacturer contracted a small data analytics company to provide value-added analytics and predictive maintenance based on data from the manufacturer’s plants.
Outsourcing data gathering & analytics	Farmers contract a small data company to obtain precision agriculture consultancy based on the images gathered by the drones of the data company.
Data reselling	A data company gathers data about mobile phone coverage through an app, and sells aggregated data to telecom operators.
Innovation space	“Industrial data spaces” have been created where traditional manufacturing companies can share their data with analytics providers in a “safe space”.
Joint venture	A large retailer made a joint venture with a data analytics company to deliver a new service to other retailers based on the retailer’s loyalty card data.
Partnership	One of the largest energy providers in the UK has partnered with data analytics company to provide analytics energy efficiency services to the energy provider customers.
Open Data	A large bank provides researchers with open access to aggregated data from the financial transactions of its customers in order to obtain new data about tourism.
Community sharing	A Greek cloud computing service allows farmers to benchmark their performance by sharing data about their farm with other farmers in the same cooperative.

b) Analysis by sector: Three patterns

There are three main patterns that could be distinguished between sectors (Table 3). In a first group, comprising aerospace, machinery and retail the most commonly adopted model consists of *outsourcing analytics*, where the traditional company, the data holder, keeps full control over how the data are used, and the subcontractor, the data company, is usually not authorised to reuse the data.

Table 3. Frequency of organizational modes per sector

	Aero	Hea	Agri	Mach	Ret	Fin	Tel	Chem	Ener	Auto	Tra	TOT
<i>Total closed</i>	3	10	10	9	9	7	9	8	6	9	9	89
Acquisition	1	1	1	1						3		7
In-house	1					2	3	3	2	3	3	17
Outsourcing data analytics		4	4	4	6	4	6	2	3	2	3	38
Outsourcing data gathering & analytics	1	5	5	4	3	1		3	1	1	3	27
<i>Total shared</i>	3			2	1	1	1	1	2			11
Data reselling							1					1
Innovation space				1								1
Joint venture				1	1	1						3
Partnership	3							1	2			6
<i>Total open</i>			1			1						2
Open data						1						1
Community sharing			1									1
<i>Grand totals</i>	6	10	11	11	10	9	10	9	8	9	9	102

A second group, comprising finance, telecommunications, chemicals, energy, automotive and transport companies, adopts in-house solutions, retaining full control of the data and providing internally the activities in big data value chain.

Finally, agriculture and health share the approach of outsourcing both data-gathering and analytics. Health is a particular case, since health data exchanges are heavily regulated and shared for research purposes (Manyicka et al., 2011). For agriculture, for instance John Deere (tractor producer) provides data gathering and analytics services to farmers, because it gather data about the farm from sensors installed in the machines. The data are formally owned by the farmer, but John Deere gathers them, stores them and can use them for performance enhancement of the machines.

c) Industry characteristics and the organizational modes selected by traditional businesses

In order to shed some light on the selection of different organizational modes by the traditional companies in the big data ecosystem, we explore whether the data intensity of a sector affects the choices of the organizational mode used. We highlight in our sample that the propensity towards data sharing is fairly consistent: companies in both the high and low data intensive sectors show a similar propensity to maintain the data in a closed mode (Table 4).

Table 4 Closed versus open modes per sector data intensity

	High data intensity sectors	Low data intensity sectors	Total
Closed	84%	90%	88%
Open	5%	2%	3%
Shared	11%	8%	9%
Total general	100%	100%	100%

?

Yet when looking at the specific modes adopted (Table 5), there are more differences between the two groups. As can be expected, in-house is much more common in high data intensive sectors (26% vs. 8%), while outsourcing data gathering is much more common in low data intensive ones (36% vs. 5%), pointing to a possible lack of data availability and need to accumulate new data in those sectors. Also, acquisition was found only in low data intensive sectors, suggesting that traditional companies are trying to reinvent themselves and acquire new data companies to create both the data and the skills needed for implementing big data. On the other hand, joint ventures and partnerships between traditional companies and data analytics companies are adopted by both groups of sectors.

Table 5 Organizational Modes per sector data intensity

	High data intensity sectors	Low data intensity sectors	Total
Closed	84%	90%	88%
<i>Acquisition</i>	0%	5%	4%
<i>In house</i>	26%	8%	13%
<i>Outsourcing data analytics</i>	53%	41%	44%
<i>Outsourcing data gathering and analytics</i>	5%	36%	28%
Shared	11%	8%	9%
<i>Data reselling</i>	5%	0%	1%
<i>Innovation space</i>	0%	2%	1%
<i>Joint venture</i>	5%	3%	4%
<i>Partnership</i>	0%	3%	3%
Open	5%	2%	3%
<i>Community sharing</i>	0%	2%	1%
<i>Open data</i>	5%	0%	1%
Total general	100%	100%	100%

2

d) Closed is the most typical mode even for the delivery of new products and services

The analysis of the business value and the mode chosen by the traditional company in the sample to share (or not) their data in the big data ecosystem adds relevant additional insights (Table 6). Closed organizational modes remain the most common across all categories of business value: productivity improvement (in 90% of the cases), segmentation (94%), improved product and service (94%), new product or service (65%). In-house (35%) and acquisitions (29%) remain the most common mode for the delivery of new products and services, while outsourcing analytics (46%) are common for boosting productivity and sales. Open (6%) and shared data (29%) are not the most used organizational modes to deliver new products or services. However, the development of new products and services relies far less than the average on closed data modes (65% vs. 87%) and more on shared modes (29% vs. 11%).

Table 6. Frequency of organizational mode per business value

	Productivity improvement	Segmentation	Improved product and service	New product or service	Total
<i>Closed</i>	90%	94%	94%	65%	87%
Acquisition	2%	0%	6%	29%	7%
In-house	8%	12%	28%	35%	17%
Outsourcing data analytics	46%	59%	28%	0%	37%
Outsourcing data gathering and analytics	34%	24%	33%	0%	26%
<i>Shared</i>	10%	6%	0%	29%	11%
Data reselling	0%	0%	0%	6%	1%
Innovation space	2%	0%	0%	0%	1%
Joint venture	2%	0%	0%	12%	3%
Partnership	6%	6%	0%	12%	6%
<i>Open</i>	0%	0%	6%	6%	2%
Community sharing	0%	0%	0%	6%	1%
Open data	0%	0%	6%	0%	1%
<i>Total general</i>	100%	100%	100%	100%	100%

Applying here the previous analysis of sector data intensity (Table 7), we appreciate that companies in low data intensive sectors focus more on productivity improvements (59%), while firms in high data intensive industries are far more likely to introduce new products and services with the adoption of big data (32%). Yet we found several cases of companies in low data intensive sectors introducing new data products and services (10%), showing that the creation of new products and services with big data is clearly applicable also to low data intensive sectors.

Table 7 Business value per sector data intensity

	High data intensity sectors	Low data intensity sectors	Total
Productivity improvement	32%	59%	53%
Segmentation	16%	14%	14%
Improved product and service	21%	17%	18%
New product or service	32%	10%	15%
Total general	100%	100%	100%

□

e) Polarization within closed and open modes

Results of the study confirm that the open data versus closed data dichotomy is insufficiently nuanced to capture the fundamental differences in traditional companies' behavior with regard to data sharing in the big data ecosystem, and that open modes are not as of today the most common modes when delivering new products and services through big data adoption. But the findings go well beyond this: they show major differences within each category (Table 8).

Table 8. Distribution of modes per business value

Value proposition Category	Improvements (Productivity improvement, segmentation, & improved product and service)	New products and services
Closed data	Outsourcing analytics	In-house analytics
	Outsourcing data gathering and analytics	Acquisitions and VC investment
Shared data	Data reselling and licensing	Joint ventures
	Open innovation Spaces	Partnership
Open data	Open data	Community sharing

In particular, within each category there is a clear distinction between modes used for improvements (productivity improvement, segmentation, improved product and services); and the modes used for delivering new products and services. For the latter, the most used organizational modes are, in the “closed” category, in-house and acquisition (29 and 35% respectively vs. 0% for the others); in the “shared” category, partnership and

joint ventures (12% each vs. 6% and 0%); and in the “open” category, community sharing (open 6% vs. 0%). Perhaps most importantly, what emerges clearly is that outsourcing, data reselling and innovation spaces are not used for delivering new products and services. In other words, across the different categories, selective and targeted forms of data sharing are used more than universal and group-based access. Companies prefer to share data through direct one-to-one contacts rather than through group-based access. This confirms the findings regarding the nascent nature of the data marketplace (Thomas & Leiponen, 2016; Carnelley, Schwenk, Cattaneo, Micheletti, & Osimo, 2016) and casts some doubt upon the OECD recommendation for nondiscriminatory access regimes (OECD, 2015), since companies seem to prefer discriminatory practices or (in the definition of Open Data Institute, 2016) named access. Last but not least, this suggests that the generation of new products and services through big data implementation requires strong institutional agreements, perhaps in order to mitigate uncertainty about the value of data.

Based on these findings, it is possible to formulate an initial conceptual framework of how traditional companies adopt different data sharing practices based on the business value they pursue and industry characteristics. Table 9 summarizes the different options adopted by the cases analyzed in the sample.

Table 9. Most common organizational modes according to business value and sector data intensity

	Productivity improvement	Sales increase	Improved product and service	New product or service
High data intensity sectors	In house Outsourcing analytics	In house Outsourcing analytics	In house Joint venture	In house
Low data intensity sectors	Outsourcing data gathering and analytics	Outsourcing analytics	Outsourcing analytics Acquisition Partnership	Acquisition Joint venture

?

This should be treated not as a recommendation for future action, but merely as a summary of current practices in our sample of early adopters in traditional industries. It is entirely possible that in the future, as companies become more accustomed to data sharing, different settings will emerge.

5. Discussion

The results of the study show that most companies in traditional sectors—the data holders—do not share data with other players in the data ecosystem (only 13 out of 102 cases), although they are more likely to share when the business value they foresee by big data implementation is developing new product and services (29% vs. 11%). The results show that great differences take place within, rather than between, the categories of open, shared and closed modes: the types of business value—productivity improvement, segmentation, improved product or service, and new product or service—are clearly associated with a specific mode of data sharing within the category. For instance, within the category of closed data, outsourcing to a data company in the ecosystem (whether for data collection and/or for including data analytics in the value chain) is only adopted for

improvement purposes (between 24% and 59% of the cases) and never for new products and services (0%). Within the category of shared modes, joint ventures and partnerships are used more when the business value sought by the company is for new product and services (in both cases 12%) than for improvements (0 to 6%).

These differences suggest that the determinants of data sharing practices do not lie only in deliberate strategic choices related to openness versus closeness, but depend on other factors. Based on the literature and the analysis carried out, in this section we formulate the following deeply interrelated hypothetical factors.

a) Uncertainty over the value of data

There is uncertainty over the value of data and how to calculate it (Thomas & Leiponen, 2016; Pantelis & Aija, 2013). It becomes challenging for data holders—traditional companies—to share data with other players in the big data ecosystem when the value of the good they could share is uncertain. From the literature we acknowledge that data previously considered non-valuable acquire value when combined with other data or analyzed in a different way (Manyika et al., 2011). Intuitively, it follows that the more unorthodox the use of the data is, the more uncertain the value. Such uncertainty might explain why the most innovative applications of big data—when the business value is foreseen in the form of new products and services—are based on in-house, acquisitions and joint venture modes, according to our sample, while transactional agreements such as outsourcing and data reselling are associated with incremental improvements (e.g. productivity improvements, improved products and services). The uncertainty over the value of data might explain why sharing data by traditional companies in the big data ecosystem is less common with regard to data than when it comes to knowledge, where OI literature has extensively documented practices of open modes for combining external and internal sources of knowledge to innovate (e.g. Bogers et al., 2014).

b) Availability of skills

Extracting value from data requires not only business skills and data analytics skills, but a peculiar *data mindset* (Mayer-Schönberger & Cukier, 2013). As has been the case for distributed process of innovation documented by OI (e.g. Christensen et al. 2005; Chesbrough & Crowther, 2006), industry characteristics—in the present study captured exclusively through sector data intensity—shed light on the selection of particular mechanisms or modes implemented by companies. Low data intensive sectors with lower data skills have to rely more on external skills and look for analytical capabilities in the big data ecosystem, typically through the outsourcing mode (adopted in 76%), while more data intensive sectors, which typically have greater data skills, can develop in-house solutions (in 26% of the cases vs. 8% in low data intensive sectors). But low data intensive sectors can also develop new products and services through acquisition, joint venture and partnering modes (adopted in 11% of the low data intensive sectors vs. 5% in high).

c) Data ownership

While it is clear that data from internal IT systems are owned by the company, there is uncertainty over the ownership of sensor-generated data (European Commission 2014). Who owns data generated by sensors installed in tractors used for agriculture, or in beacons in the shop, or in machines installed in industrial plants? Who can use the data, and who can decide who uses the data? There are no intellectual property rights over raw data, as opposed to databases or patents (Thomas & Leiponen, 2016). In fact, the greatest control over the data is attributed to the developer of the software that controls the sensor (e.g. in the tractor or in the beacon). This is why, in the case of sensor-generated data, the most common solution is the outsourcing mode for both data gathering and analytics (55% of the cases), where the data company producing the sensor has control over the data gathered for the company in the traditional sector.

d) Technical issues related to interoperability of data

Last but not least, while there is a general trend towards greater interoperability (OECD, 2015), significant effort by companies goes into preparing the data for analysis—the activity of aggregation and organization in the data value chain-, which is typically not much valued by the market (Manyika et al., 2011). This could also explain the limited uptake of data sharing practices by traditional companies in the big data ecosystem and why modes such as data reselling, or “data only” solutions where “data wrangling” costs are the main offering, are less present than modes such as outsourcing analytics that bundle data preparation with the overall data analytics service (1% versus 44%), which confirms the still incipient data market (Thomas & Leiponen, 2016; Cattaneo et al., 2016; European Commission, 2016).

6. Conclusions and future research

The present paper is a first step in an attempt to understand how companies in traditional sectors are sharing data in the big data ecosystem and, more precisely, the various mechanisms they use for sharing data with other players in the ecosystem to extract value from data across the data value chain. By addressing this question, the study fills the knowledge gap on companies’ adoption of data sharing and reuses practices and contributes to the OI literature, showing that organizational modes for knowledge sharing can be applied (and expanded) to data-sharing practices in big data ecosystems. Using the theoretical lenses of OI literature, the paper analyses the behaviour of traditional companies (data holders) in the big data ecosystem and contributes to theory development by providing a conceptual framework that associates the organizational modes that companies adopting big data in traditional sectors use to share data in the big data

ecosystem, with the business value that they foresee when adopting big data, and data-intensity of the sector.

Our findings show that sharing is less common with regard to data than when it comes to knowledge—as documented extensively by OI literature (e.g. Bogers et al., 2014)—arguably because the value of data is harder to define and the lack of clarity sometimes about data ownership. Prospective studies could undergo further empirical testing to validate these early findings and explore the current uncertainties over the value of data and how these affect the chosen organizational modes or their reasons to share or not data (Wareham, 2003). In particular, there is still limited evidence of whether those companies that are keener to share their data with other players in the big data ecosystem are also more innovative and competitive than their less forthcoming competitors. In addition, there are many differences across sectors arguably attributable to data intensity and probably other potential factors that future research could explore.

Multiple managerial implications can already be identified. Unless a company has the resources to build in-house or acquire another company, the most common solutions to create new products and services are joint ventures and partnerships, rather than market-based solutions such as data reselling. While such commercial solutions are fairly common in the context of knowledge sharing (e.g. IPR trading), they are far less common for data sharing. Market mechanisms for data sharing are little used, perhaps because it is hard to say what the data are worth.

7. References

Acha, V., and Cusmano, L. (2005) Governance and co-ordination of distributed innovation processes: patterns of R&D co-operation in the upstream petroleum industry. *Economics of Innovation and New Technology*, **14(1-2)**, 1-21.

Adner, R. (2006) Match your innovation strategy to your innovation ecosystem. *Harvard Business Review*, **84**: 98.

Akter, S. and Wamba, S.F. (2016) Big data analytics in E-commerce: a systematic review and agenda for future research. *Electronic Markets*, **26(2)**, pp.173-194.

Alavi, M. and Leidner, D. E. (2001) Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS quarterly*, 107-136.

Anderson, E. W., Fornell, C. and Mazvancheryl, S. K. (2004) Customer satisfaction and shareholder value. *Journal of marketing*, **68(4)**, 172-185.

Ansari, A. and Mela, C. F. (2003) E-customization. *Journal of marketing research*, **40(2)**, 131-145.

Autio, E., and Thomas, L.D.W. (2014) Innovation ecosystems: Implications for innovation management. In M. Dodgson, D. M. Gann and N. Phillips (Eds.), *Oxford Handbook of Innovation Management*. Oxford, UK: Oxford University Press.

Banerjee, S., Bolze, J. D., McNamara, J. M. and O'Reilly, K. T. (2011) *How Big Data can fuel bigger growth*. Accenture Outlook, (3), 28-35.

Barton, D. and Court, D. (2012) Making advanced analytics work for you. *Harvard Business Review*, **90(10)**, pp.78-83.

Basole, R. C. (2009) Visualization of interfirm relations in a converging mobile ecosystem. *Journal of information Technology*, **24(2)**, 144-159.

Bhimani, A. and Willcocks, L. (2014) Digitisation, 'Big Data' and the transformation of accounting information. *Accounting and Business Research*, **44(4)**, pp.469-490.

Bianchi, M., Cavaliere, A., Chiaroni, D., Frattini, F. and Chiesa, V. (2011) Organisational modes for Open Innovation in the bio-pharmaceutical industry: An exploratory analysis. *Technovation*, **31(1)**, pp.22-33.

Bogers, M., Zobel, A.K., Afuah, A., Almirall, E., Brunswicker, S., Dahlander, L., Frederiksen, L., Gawer, A., Gruber, M., Haefliger, S. and Hagedoorn, J., (2017) The open innovation research landscape: Established perspectives and emerging themes across different levels of analysis. *Industry and Innovation*, **24(1)**, pp.8-40.

Cattaneo, G., Glennon, M., Lifonti, R., Woodward, A., Kolding, M., and Osimo, D. (2016) *European Data Market SMART 2013 / 0063 D8 — Second Interim Report*. Retrieved from www.datalandscape.eu

Carnelley, P., Schwenk, H., Cattaneo, G., Micheletti, G., and Osimo, D. (2016) *European Data Market Europe 's Data Marketplaces – Current Status and Future Perspectives*.

Chesbrough, H. (2003a) The era of open innovation. *Sloan Management Review*, **44**, 33, 35–41.

Chesbrough, H. (2003b) *Open Innovation: the New Imperative for Creating and Profiting from Technology*. Boston: Harvard Business School Press.

Chesbrough, H.W., West, J. and Vanhaverbeke, W. (2006) *Open Innovation: Researching a New Paradigm*. Oxford: Oxford University Press.

Chesbrough, H. and Crowther, A.K. (2006) Beyond high tech: early adopters of open innovation in other industries. *R&D Management*, **36(3)**, pp.229-236.

Chesbrough, H., and M. Bogers (2014) Explicating Open Innovation: Clarifying an Emerging Paradigm for Understanding Innovation. In *New Frontiers in Open Innovation*, edited by H. Chesbrough, W. Vanhaverbeke, and J. West, 3–28. Oxford: Oxford University Press.

Chui, J. Manyika, J. Bughin, B. Brown, R. Roberts, J. Danielson, et al. (2013) *Ten IT-enabled business trends for the decade ahead: Updated research*.

Christensen, J.F., Olesen, M.H. and Kjaer, J.S. (2005) The industrial dynamics of Open Innovation—evidence from the transformation of consumer electronics. *Research Policy*, **34**, 1533–1549.

Curry, E., (2016) The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches, in: Cavanillas, J.M., Curry, E., Wahlster, W. (Eds.), *New Horizons for a DataDriven Economy*. Springer International Publishing, pp. 29–37. doi:10.1007/978-3-319-21569-3_3

Davenport, T.H. (2006) Competing on analytics. *Harvard Business Review*, **84(1)**, p.98.

Davenport, T.H. (2013) Analytics 3.0. *Harvard Business Review*, **91(12)**, pp.64-+

Demchenko, P. Membrey, P. Grosso, C. de Laat (2013) Addressing Big Data Issues in Scientific Data Infrastructure,” in *First International Symposium on Big Data and Data Analytics in Collaboration (BDDAC 2013)* Part of The 2013 Int. Conf. on Collaboration Technologies and Systems (CTS 2013), May 20-24, 2013, San Diego, California, USA.

Devaraj, S., & Kohli, A. K. (2003) Performance impacts of information technology: Is actual usage the missing link? *Management Science*, **49(3)**: 273-289.

Dretske, F. (1981) *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press

European Commission (2014) A European strategy on the data value chain. Towards a thriving data-driven economy, Communication from the commission to the European Parliament, the council, the European economic and social Committee and the committee of the regions, Brussels.

Filippov, S. (2014) Data-driven business models: powering startups in the digital age. European Digital Forum. Retrieved from:

<http://www.europeandigitalforum.eu/index.php/component/attachments/attachments?id=347&task=view>

Gandomi, A. and Haider, M. (2015) Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, **35(2)**, pp.137-144.

Gassmann, O. and Enkel, E. (2004) Towards a theory of open innovation: three core process archetypes. *Proceedings of the R&D Management Conference*, Lisbon, Portugal, July 6–9.

Gopalkrishnan, V., Steier, D., Lewis, H., Guszczka, J. and Lucker, J. (2013) *Big Data 2.0: New business strategies from big data*. Deloitte Review: 54-69.

Hammell, H. Lewis, C. Perricos, and D. Branch, (2012) Open growth: Stimulating demand for open data in the UK, in *Deloitte Analytics Briefing Note*, ed. London, UK: Deloitte, pp. 1– 12.

Hartmann, P.M., Zaki, M., Feldmann and N., Neely, A (2016) Capturing value from big data—a taxonomy of data-driven business models used by start-up firms. *International Journal of Operations & Production Management*, **36(10)**, pp.1382-1406.

Hughes, B. and Wareham, J. (2010) Knowledge arbitrage in global pharma: A synthetic view of absorptive capacity and open innovation. *R&D Management* **40 (3)**: 324-343.

Iansiti, M. and Levien, R. (2004) *The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability*: Harvard Business School Press.

Iansiti, M. and Levien, R. (2004 b) Strategy as ecology. *Harvard Business Review*, **82**: 68-78.

Krishnan, V. and Ulrich, K. T. (2001) Product development decisions: A review of the literature. *Management science*, **47(1)**, 1-21.

Lee, M.J., E. Almirall, and J. Wareham (2016) Open Data & Civic Apps: 1st Generation Failures, 2nd Generation Improvements. *Communications of the ACM*. **59(1)** pp. 82-89

Lichtenthaler, E. (2004) Organising the external technology exploitation process: current practices and future challenges. *International Journal of Technology Management*, **27(2-3)**, pp.255-271.

Lichtenthaler, U. (2005) External commercialization of knowledge: Review and research agenda. *International Journal of Management Reviews*, **7(4)**, pp.231-255.

Loebbecke, C. and Picot, A. (2015) Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *The Journal of Strategic Information Systems*, **24(3)**, pp.149-157.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H. (2011) *Big Data: The Next Frontier for Innovation, Competition, and Productivity*.

McKinsey Global Institute. Retrieved from: <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>

Mayer-Schönberger, V. and Cukier, K. (2013) *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Maxwell, J. A. (2012) *Qualitative Research Design: An Interactive Approach: An Interactive Approach*.

McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J. and Barton, D. (2012) Big data. *The management revolution. Harvard Business Review*, **90(10)**, pp.61-67.

Miller, H.G. and Mork, P. (2013) From data to decisions: a value chain for big data. *IT Professional*, **15(1)**, pp.57-59.

Mithas, S., Tafti, A. R., Bardhan, I. and Goh, J. M. (2012). *Information technology and firm profitability: mechanisms and empirical evidence*. Working paper, Robert H. Smith School of Business, University of Maryland, College Park

Moore, J. F. (1993) Predators and prey: a new ecology of competition. *Harvard Business Review*, **71**: 75-86.

Moore, J. F. (1996) *The Death of Competition: Leadership and strategy in the age of business ecosystems*. New York, NY: HarperBusiness.

OECD (2011) Exploring the Economics of Personal Data: a Survey of Methodologies for Measuring Monetary Value, *OECD Digital Economy Papers*, n. **220**

OECD (2013) Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues raised by Big Data, *OECD Digital Economy Papers*, n. **222**

OECD (2015) *Data-Driven Innovation: Big Data for Growth and Well-Being. OECD Digital Economy Papers (Vol. 215)*.

Open Data Institute (2016). Open enterprise : how three big businesses create value with open innovation, 1–25. Retrieved from <http://theodi.org/open-enterprise-big-business%5Cnhttp://theodi.org/open-enterprise->

Opresnik, D. and Taisch, M. (2015) The value of Big Data in servitization. *International Journal of Production Economics*, **165**, pp.174-184.

Pantelis, K., & Aija, L. (2013, October). Understanding the value of (big) data. In *Big Data, 2013 IEEE International Conference on* (pp. 38-42). IEEE.

Parmar, R., Mackenzie, I., Cohn, D. and Gann, D. M. (2014) The new patterns of innovation. *Harvard Business Review*, **92(1/2)**: 86-95.

Pennock, M. (2007). Digital curation: A life-cycle approach to managing and preserving usable digital information. *Library and Archives Journal*, **1**, 1–3.

- Penrose, E. T. (1959) *The Theory of the Growth of the Firm*. New York: John Wiley & Sons Inc. Penrose, E. T, 1, 1–23.
- Piccoli, G. and Pigni, F. (2013) Harvesting External Data: The Potential of Digital Data Streams. *MIS Quarterly Executive*, **12**, pp. 53– 64, 2013.
- Porter, M. E. (1985) *Competitive advantage: Creating and sustaining superior performance*. New York: Free Press. doi:10.1182/blood-2005-11-4354.
- Stefansson G. (2002) Business-to-business data sharing: a source for integration of supply chain. *International Journal of Production Economics*, **75**:135–46
- Tansley, A. G. (1935) The use and abuse of vegetational concepts and terms. *Ecology*, **16**, 284–307.
- Tether, B. S. (2002) Who co-operates for innovation, and why: an empirical analysis. *Research policy*, **31(6)**, 947-967.
- Thomas, L.D. and Leiponen, A. (2016) Big data commercialization. *IEEE Engineering Management Review*, **44(2)**, pp.74-90.
- Tuomi, I. (1999) Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. In *Systems Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on* (pp. 12-pp). IEEE.
- Ubaldi, B. (2013) Open government data: Towards empirical analysis of open government data initiatives. *OECD Working Papers on Public Governance*, **(22)**, 0_1.
- Vanhaverbeke, W., West, J., and Chesbrough, H. (2014) Surfing the New Wave of Open Innovation Research. In Henry Chesbrough, Wim Vanhaverbeke and Joel West (eds.), *Open Innovation: New Frontiers and Applications*. Oxford: Oxford University Press.
- Wamba, S.F., Akter, S., Edwards, A., Chopin, G. and Gnanzou, D. (2015) How ‘big data’ can make a big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, **165**, pp.234-246.
- Wareham, J. (2003) Information assets in interorganizational governance: exploring the property rights perspective. *IEEE Transactions on Engineering Management*. **50(3)**, pp. 337-351.
- Wareham, J., Fox, P.B., and Cano Giner, J.L. (2014) Technology ecosystem governance. *Organization Science*, **25(4)**, 1195–1215
- Weill, P. and Woerner, S.L. (2015) Thriving in an increasingly digital ecosystem. *MIT Sloan Management Review*, **56(4)**, p.27.
- West, J., and M. Bogers. (2014) Leveraging External Sources of Innovation: A Review of Research on Open Innovation. *Journal of Product Innovation Management* **31 (4)**: 814–831.

Zhu, H. and Madnick, S.E. (2009) Finding new uses for information. *MIT Sloan Management Review*, 50(4), p.18.

Zikopoulos, P., Parasuraman, K., Deutsch, T., Giles, J. and Corrigan, D. (2012) *Harness the power of big data The IBM big data platform*. McGraw Hill Professional.

8. Appendix: Code

Sectors		Aerospace
		Agriculture
		Automotive
		Chemical
		Energy
		Finance
		Health
		Machinery
		Retail
		Telecommunication
		Transport
Data sources		Internal data
		Data from sensors or physical device
		Data generated from crowdsourcing or web collaboration
		Acquired data, commercially available (e.g social media data, market data)
		Data from business partners, consumers and suppliers
		Open data
Key activity		Data gathering
		Data storage-aggregation
		Data analytics
		Data service provision
Organizational mode	Closed	In-house
		Acquisition
		Outsourcing data analytics
		Outsourcing data-gathering
	Shared	Partnership
		Joint venture
		Data reselling
		Innovation space
	Open	Community data,
		Open data
Business value		Productivity improvements
		Segmentation
		Improved product or service
		New product or service

ESADE

Ramon Llull University

Campus Barcelona - Pedralbes

Av. Pedralbes, 60-62
08034 Barcelona (España)
T. +34 932 806 162
F. +34 932 048 105

Campus Barcelona - Sant Cugat

Av. de la Torreblanca, 59
08172 Sant Cugat del Vallés
Barcelona (España)
T. +34 932 806 162
F. +34 932 048 105

Campus Madrid

Mateo Inurria, 25-27
28036 Madrid (España)
T. +34 913 597 714
F. +34 917 030 062

www.esade.edu