



HAL
open science

Big data and Smart data: two interdependent and synergistic digital policies within a virtuous data exploitation loop

Jean-Sébastien Lacam, David Salvetat

► To cite this version:

Jean-Sébastien Lacam, David Salvetat. Big data and Smart data: two interdependent and synergistic digital policies within a virtuous data exploitation loop. *Journal of High Technology Management Research*, 2021, 32 (1), pp.100406. 10.1016/j.hitech.2021.100406 . hal-03434863

HAL Id: hal-03434863

<https://hal.science/hal-03434863>

Submitted on 18 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**BIG DATA AND SMART DATA: TWO INTERDEPENDENT AND SYNERGISTIC
DIGITAL POLICIES WITHIN A VIRTUOUS DATA EXPLOITATION LOOP**

Jean-Sébastien LACAM

Associate Professor

ESSCA School of Management*

Clermont Recherche Management (CleRMa)

Phone: 06.72.08.85.24

jean-sebastien.lacam@essca.fr

ORCID ID: 0000-0002-2870-1461

David SALVETAT

Professor

ESSCA School of Management*

Phone: 06.58.23.11.62

david.salvetat@essca.fr

ORCID ID: 0000-0003-3153-1413

*1 rue Lakanal - BP 40348

49003 Angers cedex 01 - France

Abstract:

This research examines for the first time the relationship between Big data and Smart data among French automotive distributors. Many low-tech firms engage in these data policies to improve their decisions and performance through the predictive capacities of their data. A discussion emerges in the literature according to which an effective policy lies in the conversion of a mass of raw data into so-called intelligent data. In order to understand better this digital transition, we question the transformation of data policies practiced in low-tech firms through

the founding model of 3Vs (Volume, Variety and Velocity of data). First of all, this empirical study of 112 French automotive distributors develops the existing literature by proposing an original and detailed typology of the data policies practiced (Low data, Big data and Smart data). Secondly, after specifying the elements of the differences between the quantitative nature of Big data and the qualitative nature of Smart data, our results reveal and analyse for the first time the existence of their synergistic relationship. Companies transform their Big data approach into Smart data when they move from massive exploitation to intelligent exploitation of their data. The phenomenon is part of a high-end loop data exploitation. Initially, the exploitation of intelligent data can only be done by extracting a sample from a large raw data pool previously made by a Big data policy. Secondly, the organization's raw data pool is in turn enriched by the repayment of contributions made by the Smart data approach. Thus, this study develops three important ways. First off, we identify, detail and compare the current data policies of a traditional industry. Secondly, we reveal and explain the evolution of digital practices within organizations that now combine both quantitative and qualitative data exploitation. Finally, our results guide decision-makers towards the synergistic and the legitimate association of different forms of data management for better performance.

Keywords:

Big data, Smart data, volume, velocity, variety, automotive distribution.

1. Introduction

In recent years, the term “Big data” has become extremely popular thanks to the many testimonies about the impact on the economy of this innovative digital policy (Ardito *et al.*, 2019 for a review of literature). The US consulting firm Gartner Consulting (Laney, 2001) defines Big data as large data sets which, on one hand, allow managers to access new knowledge to improve their decisions and performance (Wang *et al.*, 2016; Markus, 2017) but, on the other hand, are too large, too diverse, and generated too quickly to be treated with classical information technology (IT) and organizational tools or methods (McAfee and Brynjolfsson, 2012; Günther *et al.*, 2017). Thus, the deployment of this type of digital policy requires the firm to implement an open information system to collect and process, in real time, multiple data (internal, external, structured or otherwise) (Bag *et al.*, 2020). This data is necessary for major analysis (Chen *et al.*, 2015). As soon as the firm takes on the complexity of such a digital transformation, Big data integrates almost all of its strategic and decision-making aspects (Griffin and Wright, 2015; Newell and Marabelli, 2015). The use of Big data leads to better forecasts, which can result in better decisions. On the whole, smartphones, automobiles and many other systems that contain aspects of automation are exploited to produce data streams characterized by the three-dimensional model of the "3Vs": Volume, Variety and Velocity (Laney, 2001). Automotive distribution is therefore concerned with this digital revolution (Bag *et al.*, 2020; 2021). Now, 56% of French internet users search for their new car on the internet before going to a showroom (CCM Benchmark Institut, 2015). On average, they only visit the dealership once before buying (compared to around five visits five years earlier). The way in which professionals in automotive distribution deal with connected customers, and their vehicles, transforms business relationships, affecting new and used vehicles, vehicle parts, and after-sales service (Jullien, 2017). Efficient data policy has become a major challenge for automotive distributors. Therefore, many distributors have established platforms intersecting algorithmically extracted data from information systems with various customer data in order to predict, to a finer degree, the future needs and purchases. The digital approach is evolving and

transforming in order to make Big Data “smarter” (Russom, 2011; McAfee and Brynjolfsson, 2012). The volume, variety and velocity of data does not seem to be distinctive elements of digital development (George *et al.*, 2014; Loebbecke and Picot, 2015). Currently, the challenge lies in the ability of the organization to produce smart data (Goes, 2014).

To go on about the context mentioned, this empirical study examines the transformation of a Big data policy into a Smart data policy and the existence of a synergistic relationship between them despite their distinctive nature.

The question of the company's ability to switch from Big data to Smart data is gradually being addressed by literature (George *et al.*, 2014; Bag *et al.*, 2020; Kar and Dwivedib, 2020). A debate is growing according to which “Big” is no longer the preeminent element defining Big data (Loebbecke and Picot, 2015; Huang and Huang, 2015; Li, *et al.*, 2015). The interrogations concern how “Big” data can become “Smart” through the transformation of massive and raw data, into strategic knowledge for decision-makers (George *et al.*, 2014; Bag *et al.*, 2020; Kar and Dwivedib, 2020). In this sense, some studies present smart data as an integrated subset of more massive data management (McAfee and Brynjolfsson, 2012; Wang *et al.*, 2016). Its located and specific information (extracted from a service, a site or a mission) would feed a Big data system, both global and collaborative, bringing together the contributions of all the sources of the company (Brunswicker *et al.*, 2015). However, there is currently very little knowledge and empirical studies about these management mechanisms that transforms data into a more efficient and sustainable content (Gupta and George, 2016; Shamim *et al.*, 2019). Consequently, the goal of our research is to respond to the needs to better theoretically clarify the management of this data policy transition (Ardito *et al.*, 2019; Kar and Dwivedib, 2020).

To achieve our research goal, we used a quantitative survey to collect responses from 112 French automobile distributors. Motivated by our aim on this problem of bringing together decision-makers and management scientists, the results of the study are used to propose a

theoretical framework addressing theoretically and managerially the interdependence of Big data and Smart data.

Our study makes several important contributions. First, it fills the void in the Big data literature left by prior studies which have mainly focused on the « classical » Big data practices and which have neglected the analysis diversity of data management. However, our study shows the existence of three unrelated data policies based on the founding model of 3Vs which are developed by the automobile distributors. (Low data, Big data et Smart data). Above all, Low data polities show no profit of data, nor how it is used, Big data is confirmed as a volume, a variability and a high speed processing of data (Grover, *et al.*, 2018). In contrast, our study demonstrates that Smart data used a low amount, little variety and, proven by the difference of our previous work, a variation in the speed in which the data is processed (continuously or punctually). This diversity of digital data reveals the sophistication and the recent transformations of the firms low-tech practices. A new step in digital revolution has been identified. As proof, Smart data is seen in our study as a particular advancement which succeeds only after the execution of a mass data policy. Secondly, our results help out another weak point: the knowledge of the complexity of such a process in digital transformation. Until now, the studies are based on the opposite characteristics of Big data and Smart data (Russom, 2011; Davenport and Patil, 2012; Chen *et al.*, 2015). Furthermore, our demonstration goes on even further showing the distinctive nature of the following, and underlines the relationship between the synergic and virtuous use of these polities in a two cycle step. Firstly, it is shown that a specific selection and the production of intelligent data depends on a sample of broad and raw data that was previously made up by a Big data polity. Secondly, this form of raw data is enhanced by the new contributions brought by the previous ideas of Smart data. Big data and Smart data enhanced themselves mutually through a virtuous data exploitation loop. These results reveal and explain the evolution of digital practices within organizations that now combine both quantitative and qualitative data exploitation techniques.

Following the introductory section, Section 2 presents the theoretical foundations of the hypotheses and the research models about Big data and Smart data policies. Section 3 introduces the measures and research methodology. Section 4 presents the results and data analysis. Section 5 discusses the generalizable principles/abstract ideas synthesised from our analysis. Section 6 concludes the paper by drawing lessons in data management from the findings and section 7 suggests future research, directions and limits.

2. Literature review

2.1 The 3Vs model: a characteristic of Big data policy

The ability of a firm to collect, process and analyze data from its environment effectively can lead it to better decisions and performance (McAfee and Brynjolfsson, 2012). The real value of an organization does not necessarily come from the offers the firm produces, but from its knowledge of the data it uses, which allows it to operate at a competitive advantage. In this sense, Big data has quickly become the most common digital policy due to improvements that mean that organizational management can have access to a large quantity of data (Günther *et al.*, 2017). The volume of data generated has increased exponentially in the IT era, via, for example, social media and connected objects. IT can now store and process exabytes (1 Eb = about 250 million DVDs), or even zettabytes (1 Zb = about 8 million years of recording on UHD 8K video format). Only 10 years ago, we were talking about megabytes stored on diskettes. The use of large multidimensional data sets brings new light, for example, to the nature and organization of transactions, purchase behavior, or the efficiency of a business model (George *et al.*, 2014; Chen *et al.*, 2015). In automotive distribution, increasing numbers of vehicles have connected systems that continuously transmit multiple data with a commercial

dimension to the relevant professionals, regarding maintenance, repair and equipment requirements (Bag *et al.*, 2020; 2021).

Beyond the notion of volume, Big data is also characterized by the variety of data and the velocity of their generation. Variety is explained by the opportunity offered by Big data to process data, which is both structured (from traditional databases, customer relationship management systems, spreadsheets, *etc.*) and unstructured (derived from GPS, sensors, smartphones, emails, social networks, *etc.*). Internal and external data sources are now public (e.g. governments), private (e.g. firms), ambient (e.g. trends), collective (e.g. communities of consumers) and individual (e.g. behaviors) (Baesens *et al.*, 2016). Finally, the velocity of the generation and exploitation of data has an economic impact and allows an increase in Big data efficiency with regard to policy. Whereas a firm previously dealt with a few megabytes of data in several days, it can now stream eminently greater volumes over the same period. Having real-time information allows a firm to be much more agile than its competitors. The firm's knowledge and anticipation of customer expectations, the appropriate pricing of its offerings, the adjustment of its guarantees, or even optimal inventory management can be greatly refined to stimulate the development of its distribution activities (Boyer, 2016). The resulting improvements in the offers and sales of the firm can provide a clear competitive advantage (McAfee and Brynjolfsson, 2012).

Nevertheless, the gains enabled by a Big data policy require a significant improvement in computing, cognitive and organizational capabilities (Brown-Liburd *et al.*, 2015). The data variety, format, and the accessibility of information changes mean that the firm must adapt to the growing complexity of large data. Investments (e.g. in experts, training, software and platforms) can, therefore, be important (Golia, 2013). The deployment of such a digital policy can be influenced by the activity of the firm, its size and its organizational structure (Alles, 2015). In this sense, Russom (2011) categorizes Big data policies into four levels of data exploitation and organizational engagement. The first level concerns firms that use data

intensively through analytical statistics and predictive models that guide their decisions. The second level concerns firms that rely on dedicated database management systems and develop integrated analyses that progressively influence their direction. The third level includes firms that have a distributed computer system (e.g. networks, web and telephony) that captures various data via different sources scattered around the world. The influence of data on their decisions appears less prominent compared to first- and second-level organizations. The fourth level comprises firms that have traditional tools that were designed to handle only structured data. The tools thus became obsolete model used to conceptualize Big data, to which some authors have added variability, veracity and value in order to make their analysis more relevant (Baesens *et al.*, 2016).

Introduced by the Gartner Consultancy (Laney, 2001), the 3Vs model has been improved in recent years. We now talk about the "6Vs" of Big data, although definitions vary (Chen *et al.*, 2015). The original 3Vs model still, however, allows us to characterize changes that occur in the marketplace. Therefore, three hypotheses are presented in order to identify the deployment of Big data by French automotive distributors:

Hypothesis 1a: The more a firm adopts a Big data policy, the higher the volume of the data.

Hypothesis 1b: The more a firm adopts a Big data policy, the greater the variety of the data.

Hypothesis 1c: The more a firm adopts a Big data policy, the higher the velocity of the data.

2.2 Smart data: a redefinition of the 3Vs model

While the data collection phase may appear relatively easy for a firm engaged in a Big data policy, the extraction and analysis of data in relation to meeting managerial objectives are more complex (Brown-Liburd *et al.*, 2015; Kar and Dwivedib, 2020). As a result, many firms have problems using their data to drive sales, to improve customer relationships or to optimize

the procurement process (Wang *et al.*, 2016; Bag *et al.*, 2020). However, it is the effective conduct of this process that makes a Big data policy profitable. Raw data have little value. In this sense, a high volume, variety and velocity of data do not appear to be critical elements of a Smart data policy because this policy focuses mainly on the costly and complex transformation of data into useful knowledge (George *et al.*, 2014; Loebbecke and Picot, 2015). The crucial element lies instead in the dynamic ability of the firm to extract information that will be useful and qualitative from heterogeneous datasets (Chen *et al.*, 2015). This capacity leads to the precise generation of knowledge and “smart analyses” that serve the firm's strategic aims and the choices of its decision-makers (Goes, 2014; Ardito *et al.*, 2019). Smart data is, therefore, different from Big data as conceptualized in the original 3Vs model. It does not seem necessary to exploit a large volume and variety of data to generate value for a firm. When responding to a predefined aim, data are detected and identified more efficiently by delineating a specific and minimized search domain. The firm can then develop its experience in the selection, evaluation and smart exploitation of a smaller volume of data (Davenport and Patil, 2012). Simply collecting increasing amounts of data does not mean having more knowledge (Loebbecke and Picot, 2015).

The aim of Smart data is to have specific data that address a targeted problem (Huang and Huang, 2015). At the same time, the need for such analysis could mean a slower execution velocity for Smart data compared to the context of a Big data approach. In order to produce Smart data, managers must systematically identify the questions to which they seek answers, identify the data that will answer them and their sources, and then ensure the reliability, relevance and smart handling of these data. Finally, the integration and execution of data within a firm's decision-making require the deployment of tools, methods and analytical practices that increase the time required for execution. On the one hand, the resources and skills needed for Smart data make it more qualitative than quantitative in comparison to Big data; on the other, the demands of the Smart data generation process, and then its use in the context of the firm's strategic thinking and decision-making, make a Smart data policy more time-consuming.

In this sense, three hypotheses are drawn up in order to focus on Smart data use among French automotive distributors:

Hypothesis 2a: The more a firm adopts a Smart data, the lower the volume of the data.

Hypothesis 2b: The more a firm adopts a Smart data, the less the variety of the data.

Hypothesis 2c: The more a firm adopts a Smart data, the lower the velocity of the data.

2.3 The interdependence of Smart data and Big data

Data are, above all, a raw material that can only generate gains for organizations after being processed and exploited in a way that will transform them into organizational knowledge. A firm's ability to make data "smart" and move from Big data to Smart data is essential (George *et al.*, 2014). Many automotive distributors know how to build a clientele but are unable to extract from their data knowledge about their clients' interests, the nature of their expenses, the reasons for their purchases, or the likely impact of a future of a future promotional action. More importantly for the health of the firm, many automotive distributors do not know which customers and which products are profitable and why. Thus, operating a Smart data policy is a determining factor for firms' commercial efficiency. Smart data can enable firms to build an innovative customer relationship that is stronger, more reliable and more predictive. For example, BMW concessions are associated with the manufacturer's datamining and behavioral analysis platform, in order to pre-qualify prospects (e.g. by the analysis of speech, written messages and listed practices) and then put the people targeted for concessions in contact with salespeople with profiles matching their needs (e.g. fast sales or a need for empathy). This action is based on a refined extraction of relevant, qualified and certified data from a platform that logs the customer and prospect data from all the manufacturer's information systems. Thought of as an integrated global process, the approach improves the quality and efficiency of

the firm through an original mobilization of its human resources for new support adapted to each group of customers (Bag *et al.*, 2020). In this sense, Big data appears to be the prerequisite for Smart data.

Real-time remote diagnostics, combined with preventative maintenance, is just one example of how Smart data is changing the business of automotive distribution. As part of the repair of a vehicle, some technicians consult several production or storage sites to ensure the availability of specific parts, involve virtual experts or are assisted by smart algorithms drawing similar case data from databases. Therefore, all the previously accumulated knowledge guides the diagnosis and repair of the vehicle. By gathering this specific knowledge within all its services and sites, the firm can build an integrated and interdependent system that is available for the global and smart exploitation of data (Wang *et al.*, 2016). Thus, the firm must be cross-cutting and flexible in order to foster cooperation between the units that have specialized data and the decision-makers who have to respond to problems through the relevant use of “good” data (McAfee and Brynjolfsson, 2012; Brunswicker *et al.*, 2015). The firm's digital and collaborative networks need to connect the firm with “smart” ecosystems that bring together locally produced knowledge through decentralized Smart data approaches. The interaction of these multiple actors thus creates a set of knowledge, research opportunities and application solutions (Brunswicker *et al.*, 2015). This model manages several “smart” data flows located at the micro level and in the different contexts and environments of the firm. In this sense, Smart data's decentralized initiatives are part of a firm's Big data policy when the firm captures and consolidates its “smart” data into an open integration platform. Smart data is, therefore, a subset of Big data. It is not the size of a centralized and generalized Big data policy that generates wealth in a firm, but the value of the specialized data produced by its decentralized units (e.g. services and sites). After the data collection, the information and knowledge are consolidated by other additive or complementary elements, then prepared to be exploited instantly by users or deferred for later use. Big data that aims to improve management decision-making (Newell and Marabelli, 2015) thus capitalizes on an integrated Smart data approach. Big data and Smart

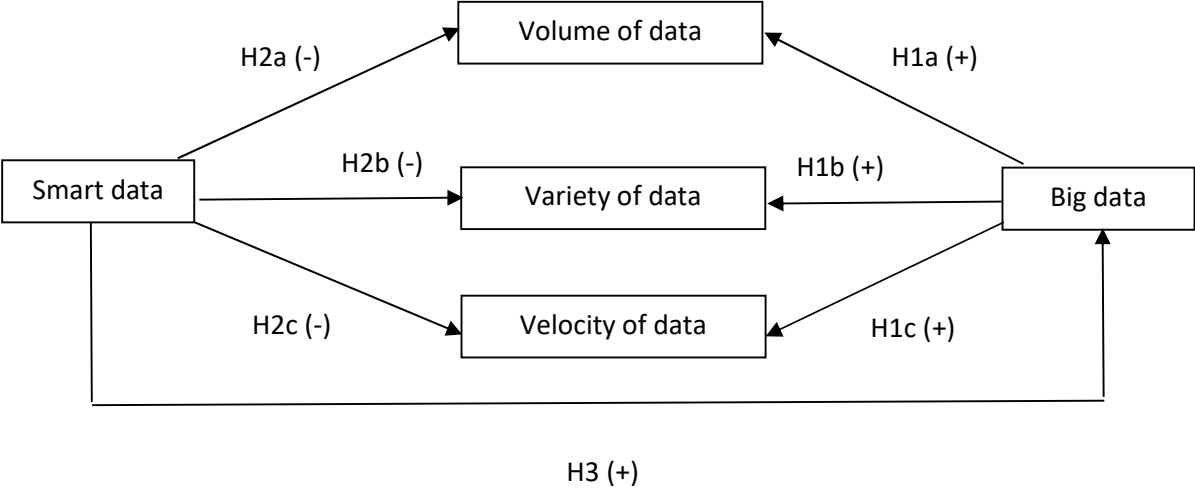
data are, therefore, interdependent policies within a single technical, organizational and smart system.

Thus, we assume the existence of a virtuous loop that would bring synergy between Smart data and Big data within the firm. A final hypothesis is, therefore, generated:

Hypothesis 3: The more a firm is engaged in Smart data, the more it is engaged in Big data.

Finally, our theoretical framework is presented in Figure 1:

Figure 1. Theoretical framework



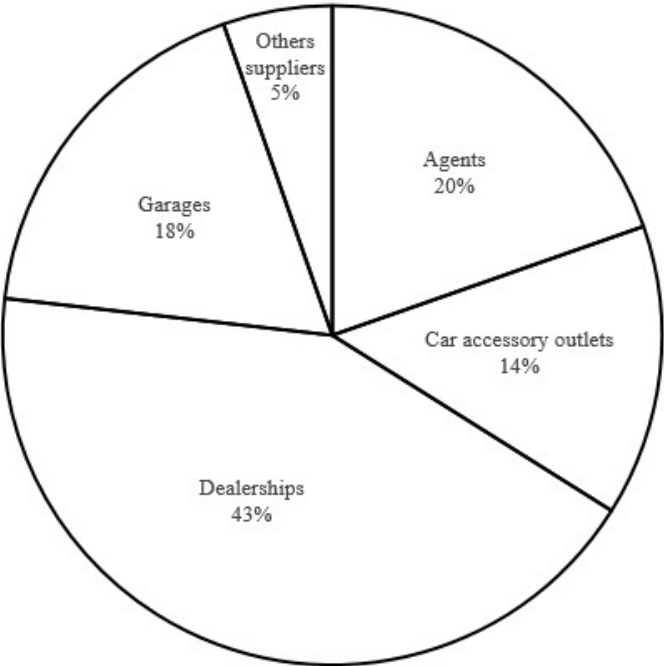
3. Methodology

3.1 Field and sampling

Our empirical study focuses on policies of digital data management among French automotive distributors. This distribution activity includes over 31,000 firms and nearly

150,000 employees engaged in the sale of new and used vehicles, as well as the maintenance and sale of spare parts (ANFA, 2016). Three types of player are identified in national distribution: manufacturers (via their own points of sale and their internal services); automotive distributors and secondary network garages that are in partnership with manufacturers; and outside network intermediaries (e.g. large retailers, authorized representatives and accessories shops) that buy from automotive distributors, the foreign subsidiaries of manufacturers, etc. (Xerfi, 2016). Our research questionnaire (available on request) addresses a sample of 1,600 of these distributors drawn on the basis of equal probabilities. A website was designed for the investigation to collect responses from the participants. After analysis (some questionnaires were not completed and some were received beyond the deadline), 112 firms were selected (see Figure 2). Based on the criteria of the European Commission (2003/361/EC), 33% of our sample are small enterprises (fewer than 50 employees and a turnover of between 2 and 10 million euros) and 67% are medium-sized enterprises (50 to 250 employees and a turnover of between 10 and 50 million euros).

Figure 2. Distribution of the selected 112 firms by category of firm and data policy



	No data	Big data	Smart data
<i>Agents</i>	33%	17%	20%
<i>Car accessory outlets</i>	17%	16%	8%
<i>Dealerships</i>	25%	43%	52%
<i>Garages</i>	25%	17%	16%
<i>Other suppliers</i>	0%	7%	4%

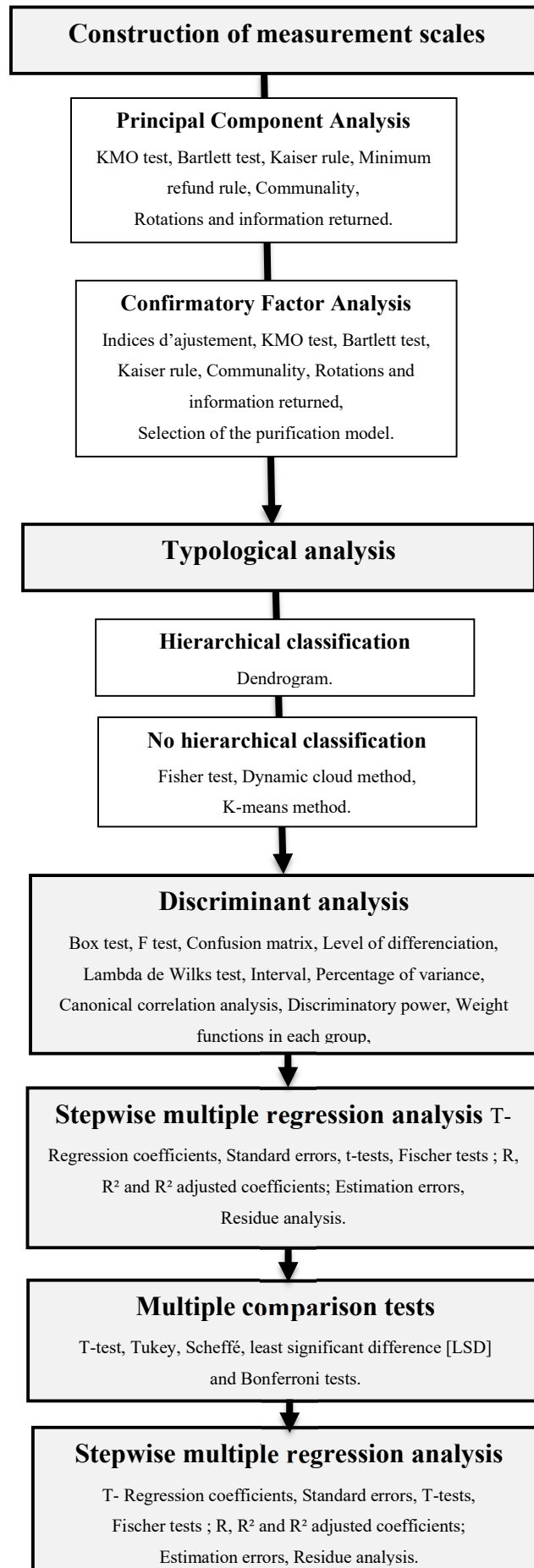
3.2 Data processing

First, the measurement scales were constructed using principal component analysis (PCA) and then validated through confirmatory factor analysis (CFA). The thresholds of acceptability of the mobilized indicators (Cronbach's alpha, absolute indices, *etc.*) and their results are presented in the appendices (see Table 1).

Second, the participating firms were divided into groups, the profiles of which diverge according to a non-hierarchical classification (the k-means clustering method). Our intention was to establish a more local optimum for the quality of the classification than a global one. Euclidean distance and the factorial scores from the PCA or CFA carried out proximity measurement in order to limit the existence of relationships between variables. Based on the factor scores, variance analysis (ANOVA) was used to validate the classification. The explanatory powers of the factorial axes (F test) and their significance were also calculated. The results of this classification were then validated by discriminant analysis. The barycenters of the groups were also determined. The study of the three macro variables (Volume, Variety and Velocity) characterizing digital data management allowed us to identify three groups of firms related to two discriminate functions. Multiple comparison tests (MCTs) (t-test, Tukey, Scheffé, least significant difference [LSD], and Bonferroni tests) differentiated the involvement of groups from the three macro variables.

Third, stepwise multiple regression analysis was conducted by considering the three groups of firms identified. This analysis determined the explanatory variables in the dependent macro variables (Velocity, Volume and Variety) to eliminate the collinearity risk (Foucart, 2000)

Figure 3. Research methodology flow chart



3.3 Macro variables and variables

Our study presents a panorama of digital data management policies. Three macro variables and dependent variables (*Velocity*, *Variety* and *Volume* of data) and 14 independent variables allowed the identification of practices and their impact on the activity of French automotive distributors. The reliability, quality and discriminating validity of these scales were validated during the PCA and CFA (see Table 1).

3.3.1 Macro and dependent variables: *Volume*, *Variety* and *Velocity* of data

Friedman and Marley (2015) and Boyer (2016) inspired the items for these three macro variables. As a reminder, these same macro variables were used to create three groups of firms according to the level of their Big data activity (cluster analysis).

First, the macro variable *Volume* of data measures the quantity of data collected and produced by the firm. This macro variable is composed of three variables: a low volume (weak perceived quantity of annual data), a standard volume, and a high volume.

Second, the variable *Variety* of data measures the diversity of data (e.g. texts, pictures and videos) as well as how current they are (old data, recent data).

Third, the macro variable *Velocity* of data measures the rate of data acquisition (continuous/discontinuous) and the speed of the resulting decision-making (delayed-time or real-time). This macro variable is composed of two variables: a low velocity of data (delayed-time decision-making and discontinuous data flow) and a high velocity of data (real-time and continuous data flow).

Table 1: Scales: PCA, CFA and acceptance thresholds

					AI					RI					PI
Macro variables	Variables	RM	REI	α	χ^2	GFI	AGFI	RMR	RMSEA	NFI	RFI	CFI	IFI	TLI	χ^2/ddf
					The + ↓	≥ 0.9		≈ 0	< 0.09	≥ 0.9					< 5
Velocity	Low	CFA	0.333	0.75	9.1	0.974	9.31	0.068	0.036	0.963	0.931	0.995	0.995	0.991	1.137
	High	CFA	0.313	0.827											
Variety		PCA	0.658	0.73											
Volume	Low	CFA	0.253	0.845	4.6	0.987	0.953	0.042	0.081	0.989	0.972	0.99	0.989	0.958	0.766
	Standard	CFA	0.279	0.961											
	High	CFA	0.241	0.823											
Business sources	Institutional	CFA	0.401	0.687	1.197	0.995	0.947	0.029	0.042	0.986	0.919	0.998	0.998	0.986	1.197
	Horizontal	CFA	0.359	0.603											
Individual sources	Personal	CFA	0.405	0.741	1	0.997	0.987	0.018	0.001	0.991	0.978	1	1.029	1.076	0.2
	Professional	CFA	0.273	0.5											
Customer relationships	Quality	PCA	0.715	0.866											
	Delivery	CFA	0.641	0.809	2.4	0.989	0.943	0.026	0.045	0.986	0.959	0.997	0.997	0.992	1.2
	Track	CFA	0.696	0.854	0.1	1	0.999	0.003	0.001	1	0.999	1	1.009	1.026	0.05
	Customer satisfaction	PCA	0.878	0.89											
Data utilization	Flexibility	PCA	0.586	0.822											
	Cost control	CFA	0.673	0.833	2.8	0.988	0.94	0.018	0.061	0.988	0.964	0.996	0.996	0.989	1.4
	Human resources	PCA	0.729	0.907											
Innovation		CFA	0.641	0.801	1.9	0.992	0.96	0.017	0.001	0.991	0.972	1	1.001	1.002	0.95
Development	Offers	PCA	0.749	0.787											
	Sales	CFA	0.713	0.863	0.474	0.998	0.99	0.007	0.001	0.998	0.995	1	1.006	1.018	0.237

Note: α = Cronbach's alpha; AGFI = Adjusted goodness-of-fit index; AI = Absolute indices; CFA = Confirmatory factor analysis; CFI = Comparative fit index; $\chi^2 = Chi^2$; $\chi^2/ddf = Chi^2$ by the degree of freedom; GFI = Goodness-of-fit index; IFI = Incremental fit index; NFI = Normed fit index; PCA = Principal component analysis; PI = Parsimony indices; REI = Return information; RFI = Relative fit index; RI = Relative indices; RM = Retained methods; RMR = Root mean residual; RMSEA = Root mean square error of approximation; TLI = Tucker-Lewis index.

3.3.2 Macro and independent variables: Sources, Customer relationships, Data utilization, Innovation and Development

The macro variables *Business sources* and *Individual sources* measures data collection channels. These macro variables are composed of four variables: the institutional sources (professional organizations, public institutions, *etc.*) and horizontal sources (competitors and allies) of a firm, and then the personal sources (friends, family, *etc.*) and professional sources (associations, unions, *etc.*) of individuals working in the firm. These macro variables are based on items proposed by George et al. (2014) and Boyer (2016).

The macro variable *Customer relationships* measures the involvement of data in the links and exchanges built between a firm and its customers. This macro variable is composed of four variables: quality (confidence, responses to needs, *etc.*), delivery (respect, extended time limits, *etc.*), track (complaints, customer involvement, assistance, *etc.*) and customer satisfaction (measuring tools, reduction of customer complaints, *etc.*). These items are based on an article by Brulhart and Moncef (2010).

The macro variable *Data utilization* measures the use made by a firm of the data it collects. This macro variable is composed of three variables: flexibility (changes caused by data relating to the means of production, offers, distribution, *etc.*), cost control (improvement of predictions, cost reduction, *etc.*) and human resources (involvement of employees, motivation, participation, *etc.*). The items are based mainly on suggestions from Xiang (2013).

The variable *Innovation* measures the assistance provided by the collected data to the innovation of the firm (products and methods). Ritala (2012) was the main inspiration for this item.

Finally, the macro variable *Development* measures the consequences of a firm taking data into account in its economic expansion policies. This macro variable is composed of two variables: offers (improvement of offers, reduction in the return rate, *etc.*) and sales (demand stimulation, confrontation with competitors, *etc.*). Brulhart and Moncef (2010) also provided the basis for these items.

3.3.3 Control variables

In line with previous research, we mobilized four control variables that are likely to affect the management of big data policy. Control variables were taken from the Diane database: The *Size of the firm* (staff), the firm's *Turnover*, its *Total assets* and its *Age*.

4. Findings

4.2 Three groups of firms and two discriminating functions identified

Typological analysis was carried out via the macro variables corresponding to the *Volume*, *Velocity* and *Variety* of data. The association between these variables allows the nature of the digital data management policies established by French automotive distributors to be studied. The barycenter scores are presented in parentheses (G) below.

The low data group was composed of 12 small-sized firms, characterized by a data velocity that is neither low ($G = -0.545$) nor high ($G = -1.643$). The data variety and data volume are low ($G = -0.411$ and $G = 0.546$, respectively).

The Big data group contained 75 medium-sized firms characterized by a high data velocity ($G = 0.221$). The data variety and data volume are high ($G = 0.336$ and $G = 0.152$, respectively).

Twenty-five small-sized firms were contained in the Smart data group and are characterized by a high data velocity ($G = 0.127$) but especially a low velocity ($G = 0.326$). The data variety and data volume are low ($G = -0.811$ and $G = 1.193$, respectively).

The discriminant power of our classification is 96.88%. Fisher and Wilks' lambda tests indicate that the macro variables of the *Velocity*, *Volume* and *Variety* of data allow for better discrimination between firms. Discriminant analysis allowed us to identify two discriminating functions (with a discriminate power of 87.20%). The *Extended data exploitation* function shows positive scores for the management of a high data volume and a high data variety. This function specifically characterizes the Big data group of firms. In contrast, the *Smart data exploitation* function presents positive scores for a high velocity of reaction to a mainly low data volume. This function characterizes the Smart data group of firms.

4.3 Characteristics of digital data management according to their Volume, Variety and Velocity of data

The analysis of centroids, analysis of variance (ANOVA) and Multiple comparison tests (MCT: T-tests, Tukey where n is unequal, Scheffé, Least Significant Difference and Bonferroni) focus on the particularities of dependent macro variables in order to identify, for each group of firms, the characteristics of data policies deployed. The barycenter symbols (G) and the significances of the Fisher's, Student's t-tests and MCTs (p) are shown in parentheses below.

4.3.1 Macro variable *Volume of data*

Concerning the volume of data, the barycenter analysis indicates that the Low data group collects and produces data at a lower volume ($G > 0$) compared to the other groups. The Big data group collects and produces data of a standard and high volume ($G > 0$). The Smart data group essentially collects and produces data to a low and standard volume and, to a lesser extent, a high volume ($G > 0$) compared to the other groups. The Fisher's tests show that there is a significant difference between the three groups ($p < 0.001$) with regard to the three types of volume (low, standard and high).

As part of the MCT, the Low data group differs from the other groups, which have a data management policy (the Big data and Smart data groups) regarding all types of volume ($p < 0.001$). The Smart data group is also distinguished from the Big data group with regard to the low volume of their data ($p < 0.001$). When we put together the Big data and Smart data groups, the Student's t-tests indicate that this new is distinguished from the Low data group for all types of volume ($p < 0.001$). Thus, the more a firm manages data "massively" (Big data), the greater the volume of data collected and produced. In contrast, the more a firm manages data "smartly" (Big data), the more the volume of the data collected and produced is low to standard.

To complete this analysis, we focus on the MCT of items related to each of the variables. Concerning the low volume of data, the item "*According to you, your firm collects a 'low' volume of data per year (one or several CD-ROMs, USB storage keys, ...)*." indicates that the Big data group differs from the Low data and Smart data groups ($p < 0.001$). The item "*According to you, your firm produces a 'low' volume of data per year (one or several CD-ROMs, USB storage keys, ...)*." indicates that the Low data group differs from the Big data ($p < 0.05$) and Smart data ($p < 0.001$) groups. The Smart data group can also be distinguished from

the Big data group for this item ($p < 0.001$). When we put together firms that practice a policy of digital data management, the Student's t-tests indicate that this new set is distinguished ($p > 0.01$) from the Low data group for the first item.

Concerning the standard volume of data, the items “*According to you, your firm collects a ‘standard’ volume of data per year (one or more hard disks for storage)*” and “*According to you, your firm produces a ‘standard’ volume of data per year (one or more hard disks for storage)*” indicate that the Low data group differs from the Big data and Smart data groups ($p < 0.001$). When we put together firms that practice a policy of digital data management, the Student's t-tests indicate that this new set is distinguished from the Low data group for both items ($p > 0.001$).

Concerning the high volume of data, the item “*According to you, your firm collects a ‘high’ volume of data per year (one or more servers of storage)*” distinguishes the Low data and Big data groups ($p < 0.01$). The item “*According to you, your firm produces a ‘high’ volume of data per year (one or more servers of storage)*” indicates that the Low data group differs from the Big data and Smart data groups ($p < 0.001$). When we put together firms that practice a policy of digital data management, the Student's t-tests indicate that this new set is distinguished from the Low data group for both items ($p > 0.01$).

Thus, firms that have a Low data management policy are characterized by a low collection and low production of data. Firms that practice Big data are only focused on a high collection and high production of data. Finally, firms that practice Smart data are centered on a low collection and low production of data. Therefore, the passage of a firm from a Low data to a Big data form is marked by the exploitation of a large volume of data. In contrast, the passage of a firm from a Big data to a Smart data form is marked by the exploitation of a low volume of data.

4.3.2 Variable Variety of data

Concerning the variety of data, the barycenter analysis indicates that the Big data group collects the most varied data ($G > 0$) compared to the other groups. The Smart data group collects the least varied data ($G < 0$), although the Low data group also collects a low level of varied data ($G < 0$). Fisher's tests show that there is a significant difference between the three groups with regard to the variety of data ($p < 0.001$). As part of the MCT, the Low data group differs from the Big data group of firms ($p < 0.05$) regarding the variety of data. The Big data group is distinguished from the Smart data group ($p < 0.001$). Thus, the more a firm manages "massive" data (the Big data group), the greater the variety of the data collected. The more a firm manages data "smartly", the lower the variety of the data collected.

To complete this analysis, we focus on the MCT of items related to each of the variables. The item "*Your firm uses several data types (encrypted data, texts, pictures, videos, ...).*" indicates that the Big data group differs from the Smart data group ($p < 0.001$). The item "*Your firm uses old data (more than one year, ...).*" indicates that the Big data group differs from both the Low data ($p < 0.05$) and Smart data ($p < 0.001$) groups. The item "*Your firm uses recent data (less than one year, ...).*" indicates that the Big data group differs from the Smart data group ($p < 0.001$). When we put together firms that practice a policy of digital data management, the Student's t-tests indicate that this new set cannot be distinguished from the Low data group for any of the items.

Thus, firms that practice Big data exploit a high variety of data, while firms that practice Smart data exploit a low variety of data (more targeted).

4.3.3 Macro variable Velocity of data

The barycenter analysis indicates that the Low data group reacts least to data compared to the other groups. In contrast, the Big data group reacts the most quickly ($G > 0$). The Smart data group displays the most delayed data exploitation ($G > 0$). Finally, the Fisher's tests show that there is a significant difference between the three groups regarding delayed data exploitation ($p < 0.05$) and instant data exploitation ($p < 0.001$).

As part of the MCT analysis, the Low data group is distinguished from the Smart data group regarding delayed data exploitation ($p < 0.05$). When we put together firms that practice a policy of digital data management (i.e. the Big data and Smart data groups), the Student's t-test indicates that this new set is not distinguished ($p > 0.05$) from the Low data group. However, all the groups can be distinguished from each other via MCT in relation to a high velocity of data exploitation ($p < 0.001$). Finally, when we again put together firms that practice a policy of digital data management, the Student's t-test indicates that this new set can be distinguished ($p < 0.001$) from the Low data group. Thus, the more a firm practices a policy of digital data management, the more immediate (faster) its data exploitation. However, the more a firm practices a specific approach to Smart data, the more delayed (slower) its data exploitation.

To complete this analysis, we focus on the MCT for items related to each of the variables. Concerning the low velocity of data, only the item "*Your firm is informed on the topics that interest it (trends, actors, etc.) with a time delay (after a certain time...)*" distinguishes the Low data group from the Smart data group ($p < 0.05$). When we put together firms that practice a policy of digital data management (the Big data and Smart data groups), the Student's t-test indicates that this new set is indistinguishable from the Low data group for all the items ($p < 0.05$).

Concerning the high velocity of data, all the items, “*Your firm is informed on the topics that interest it (trends, actors, etc.) in real time (instantly...)*.”; “*Therefore, your data help you to make decisions in streaming (continually...)*”; and “*Therefore, your data help you to make decisions in real time (instantly...)*.”, distinguished all the groups ($p < 0.001$). When we put together firms that practice a policy of digital data management (the Big data and Smart data groups), the Student's t-tests indicate that this new set is distinguished ($p < 0.001$) from the Low data group for all items.

For example, firms that practice a Smart data approach react both instantaneously and with a delay to their data. In contrast, firms in the Big data group react only in an instantaneous manner. Firms in the Low data group react only after a long time (showing the most negative barycenter that moderates their reaction). On the one hand, the passage of a firm from a Low data to a Big data form is marked by the immediacy of the data. On the other, the passage of a firm from a Big data to a Smart data form is marked by the addition of capacity to defer its use of data.

4.4 Determinants of the Volume, Variety and Velocity of data

We used a stepwise regression model to examine the dependent variables for each group to examine what characteristics of data determines their data policy (see Table 2). By way of simplification, we only present the significant results for the Big data and Smart data groups in this section. To sum up, a reduction in the *Velocity* of an organization's response to data is associated with greater *Variety* of data for the big data group and for the smart data group. An increase in a standard amount of data is also associated with a slower reaction rate for the Smart data group. However, a high *Volume* of data drives an increase in the reaction rate of a company

for the big data group. Thus, the *Velocity* of a firm's response to data is determined by the *Variety* and/or the *Volume* of the data.

Finally, we note that the control variables relating to the *Size of a firm*, its *Turnover* and its *Total assets* have no causal relationship with the *Velocity*, *Volume* and *Variety* of the firm's data.

Table 2. Stepwise regression analysis

Model 1			
Dependent variable: Low velocity			
	Low data	Big data	Smart data
Offers	0.769 (*)		
Individual personal sources	-0.494 (**)		
Variety		0.342 (**)	0.533 (**)
High velocity		-0.300 (**)	
Institutional business sources			0.351 (**)
<i>R/R²/R² adj.</i>	<i>0.868/0.753/0.698</i>	<i>0.423/0.179/0.156</i>	<i>0.713/0.508/0.464</i>
<i>Fisher tests</i>	<i>13.741 (**)</i>	<i>7.848 (*)</i>	<i>11.368 (*)</i>

(*) = $p < 0.001$; (**) = $p < 0.01$; (***) = $p < 0.05$

Model 2			
Dependent variable: High velocity			
	Low data	Big data	Smart data
Low velocity	1.273 (*)	-0.288 (**)	
High volume	0.549 (*)	0.251 (**)	
Delivery	-0.245 (***)		
<i>R/R²/R² adj.</i>	<i>0.978/0.957/0.933</i>	<i>0.355/0.126/0.102</i>	/
<i>Fisher tests</i>	<i>39.089 (*)</i>	<i>5.195 (**)</i>	/

(*) = $p < 0.001$; (**) = $p < 0.01$; (***) = $p < 0.05$

Model 3			
Dependent variable: Variety			
	Low data	Big data	Smart data
Quality	1.097 (*)		0.446 (**)
Delivery	-0.601 (**)		
Low velocity		0.313 (**)	0.431 (***)
Individual professional sources		-0.238 (***)	
Horizontal business sources		0.237 (***)	
<i>R/R²/R² adj.</i>	<i>0.937/0.877/0.85</i>	<i>0.446/0.199/0.165</i>	<i>0.744/0.554/0.514</i>
<i>Fisher tests</i>	<i>32.198 (*)</i>	<i>5.888 (*)</i>	<i>13.676 (*)</i>

(*) = $p < 0.001$; (**) = $p < 0.01$; (***) = $p < 0.05$

Model 4			
Dependent variable: Low volume			
	Low data	Big data	Smart data
Individual personal sources	-0.767 (**)		
High volume		0.249 (***)	
Delivery			0.529 (**)
Horizontal business sources			-0.357 (***)
Institutional business sources			0.348 (***)
<i>R/R²/R² adj.</i>	0.767/0.588/0.546	0.249/0.062/0.049	0.702/0.493/0.447
<i>Fisher tests</i>	14.251 (**)	4.815 (***)	10.508 (*)

(*) = $p < 0.001$; (**) = $p < 0.01$; (***) = $p < 0.05$

Model 5			
Dependent variable: Standard volume			
	Low data	Big data	Smart data
High volume		0.110 (*)	
Low velocity			0.214 (***)
<i>R/R²/R² adj.</i>		0.388/0.151/0.139	0.436/0.190/0.155
<i>Fisher tests</i>		12.976 (**)	5.394 (***)

(*) = $p < 0.001$; (**) = $p < 0.01$; (***) = $p < 0.05$

Model 6			
Dependent variable: High volume			
	Low data	Big data	Smart data
Quality	1.103 (*)		
Low velocity	-0.808 (*)		
Standard volume		-0.372 (*)	
Low volume		0.266 (***)	0.472 (***)
Institutional business sources		0.259 (***)	
<i>R/R²/R² adj.</i>	0.921/0.849/0.815	0.517/0.267/0.236	0.472/0.223/0.189
<i>Fisher tests</i>	25.293 (*)	8.624 (*)	6.600 (***)

(*) = $p < 0.001$; (**) = $p < 0.01$; (***) = $p < 0.05$

5. Discussion

Despite their possible distinctive natures, is there an interdependent relationship between Big data and Smart data in the operation of the data of a firm? French car distributors differentiate themselves through the deployment of digital policies in each of the three aspects of Big data (Volume, Variety and Velocity of data). The existence of three groups of car dealers differentiated according to their level of data management shows the following plurality of practices: firms that practice a Low data policy and do not practice any form of data management; firms that practice a Big data policy; and firms that practice a Smart data policy. We then focused our analysis on the second and third groups of firms.

5.1 Theoretical implications

First, our results are consistent with our previous work in which volume (Grover, *et al.*, 2018), variability (Gandomi and Haider, 2015) and velocity of data (Pigni, *et al.*, 2016) must be high for a policy to be characterized as Big data. The validation of our hypothesis 1a indicates that the volume of data must be high for Big data to be a feature of data storage and access (Cyganek, *et al.*, 2016; Grover, *et al.*, 2018). The data are cumulative: new features are added to old ones to improve the processing, the performance and decision-making. The volume of data is important when it reaches a certain computational limit (Alexander, *et al.*, 2015) or when the conventional tools for research, collection, storage, analysis, *etc* of these data are considered out of date (Holden, 2016). Our hypothesis 1b is validated and confirms that Big data also concerns various data that is generated from multiple sources (Alexander, *et al.*, 2015). The

rapid accumulation of data in different forms and many origins has led to a growing interest in the analysis of large data (Chiang, *et al.*, 2018). The variety of sources (Web, social media, CRM, *etc.*) has led to a rapid expansion of the diversity of data available to firms (Agarwal and Dhar, 2014). This data comes from both internal sources and those external to the company (Grover, *et al.*, 2018) and have several formats: such as emails, web browsing, maps, *etc.* (Zhao, *et al.*, 2014). Finally, our hypothesis 1c is validated and confirms that data's velocity is typical of Big data. The data streams are continuous, permanent, in real time and cumulative (Alexander, *et al.*, 2015; Grover, *et al.*, 2018). Constantly generated, data evolves rapidly over time (Cyganeck, *et al.*, 2016). The high velocity shortens the decision cycle (Pigni, *et al.*, 2016).

Secondly, previous work indicates that a Smart data policy presents a volume, a variability and a velocity of the data which must be low (Strîmbei, 2012; George, *et al.*, 2014; Li, *et al.*, 2015). Our results corroborate in part the literature. They suggest firms that practice Smart data exploit instantaneously or by deferring a low volume of data with little variation. The validation of hypotheses 2a and 2b confirms that the volume and the variety of data are low in this type of policy (George, *et al.*, 2014). The company develops its experience in the selection, evaluation and intelligent exploitation of a smaller volume and variety of data (Davenport and Patil, 2012). Simply collecting increasing amounts of data does not mean having more knowledge (Loebbecke and Picot, 2015). However, the partial validation of our hypothesis 2c reveals that the exploitation of intelligent data is delayed in time or instantaneous. There is therefore a variation in their velocity and not a systematic decrease as some previous work demonstrates (Li, *et al.*, 2015). Smart data is an approach that mobilizes specific data to a particular problem, while Big data is a global, general approach. Smart data is also differentiated by the capacity to be informed continuously or punctually. Then, used to make instant or delayed decisions. This management method places the organization in a variable temporality. High volume, variety and velocity do not, therefore, appear to be the critical

elements of a Smart data policy since it focuses mainly on transforming data into useful knowledge rather than the crucial element that lies in the firm's ability to extract qualitative information from heterogeneous masses of data that will be useful to this system (Chen *et al.*, 2015). This ability generates knowledge to support strategic goals and decision making (Goes, 2014). At the same time, the need for such an analysis can result in variable execution speed for smart data compared to the context of a big data approach. In order to produce smart data, managers must systematically identify the questions to which they are looking for answers, identify the data that will answer them as well as their sources, then ensure reliability, relevance and intelligent processing of this data. Finally, integrating and executing data into a firm's decision making can require the deployment of analytical tools, methods, and practices that increase or decrease in time.

At last, the partial validation of our hypothesis 3 confirms, for the first time, the interdependence of Big data and of Smart in the generation of data of this data in the company. Traditionally, the work dissociates these two polities of which their characteristics are distinguished (Russom, 2011 ; Huang and Huang, 2015). Although, this still questions the capability of the firm to go from Big data to Smart data to improve the exploitation of the information that is approached by literature (George *et al.*, 2014 ; Bag *et al.*, 2020; Kar and Dwivedib, 2020). To go on, some contributors present the steps to Big data as a system that is integrated to an organisation that works thanks to the world-wide coordination of multiple sources of data attached to a specific website, service, or to a particular activity (McAfee et Brynjolfsson, 2012 ; Wang *et al.*, 2016). Big data is defined as a collaborative polity which the effectiveness depends on an analyse, whole date that is specific and spread (Brunswicker *et al.*, 2015). This would make Smart data a subunit of Big data. If our results confirmed this interdependence between the exploitation of mass data and the production of smart data, they enhance and could question these ideas in two parts. First of all, the relationship between these

two types of data polities are reversed for French automotive distributors. Big data is presented as a massive and varied data storage originally made by firm, and after, exploited by Smart data. Smart data draws from Big data (and not the other way round) to be able to establish a specific selection of data needed to produce more, that is focused on being targeted and more sophisticated. Big data is a necessity before obtaining Smart data. Secondly, this relationship between both of these data policies data are linked into a virtuous data exploitation loop. After having previously drawn in a massive data storage, Smart data then enhances the Big data storage so the organisation can benefit from processing the data on a larger scale, and to seek the more diversity. There is no possible substitution of Smart data to Big data, but, more of a simultaneous and synergistic connection of these polities, who are part of a digital transformation movement conducted by low-tech organisations. Therefore, our study contributes back to theory building in data management following the request to better theoretically clarify the transition of data policy (Ardito *et al.*, 2019; Kar and Dwivedib, 2020).

First of all, our research confirms the democratisation and the diversity of data management in small low-tech companies.

The exploitation of data is becoming more and more popular amongst the PME, who are becoming aware of how necessary the investment is worth working on and its advantages to become more successful (Donnelly *et al.*, 2015). This is shown throughout the three types of data management identifications used by automotive companies, our study is strengthened assuring that smaller organisations are also looking to extract the value of their received data by creating a new digital managing model (George *et al.*, 2014; Gandomi and Haider, 2015; Günther *et al.*, 2017; Schwab and Zhang, 2018). More precisely, our results enhance the knowledge by exploring the details and characteristics of a certain policy that remain unknown by literature: Smart data. On one hand, we confirm that it remains different from the traditional

one, with the management of a low volume and almost no variety of Big data (George, *et al.*, 2014; Li, *et al.*, 2015). On the other hand, we have demonstrated for the first time, the high speed of data exploitation thanks to Smart data which distinguishes itself through an instant exploitation system or is differentiated over a period of time based on what strategy is used. In contrast, of a Big data generalist, our results highlight the sophistication and the maturity of Smart data by showing its capacity to adjust the management of particular content linked to a previous targeted problem. Secondly, our research brings a new spotlight on our previous work showing the complexity of data analysis, enabling and achieving the tasks to get more answers on the previous management goals (Brown-Liburd *et al.*, 2015). The three V model is much more than today's unique data model (Davenport et Patil, 2012). Besides this crucial element, being able to collect useful qualitative information from massive heterogeneous data (Chen *et al.*, 2015; Bag *et al.*, 2020). For example, our study specifies that a large variety of data slows down the speed of information and the decisions of an organisation, no matter what the digital policy is. The use of massive heterogeneous data represents a struggle in reaching their goals. To finish, if our results confirmed the literature of this Big data evolution, in becoming a smarter digital policy (Russom, 2011; McAfee et Brynjolfsson, 2012), this brings new knowledge to the Big data's transformation and the process of it becoming Smart throughout the discovery of their synergistic relationship. Our study suggests, for the first time, a fused vision of both of these policies, contradicting the traditional relationship (Russom, 2011; Chen *et al.*, 2015). Instead of an opposite phenomenon, or even a substitutional one, the quantitative side of Big data and the qualitative of Smart data are complementary and codepend.

5.2 Practical implications

Our results have significant implications for digital data management.

First, managers need to recognize the importance of the opportunities (and limitations) presented by varying degrees of data exploitation. While data processing has proven to be associated with faster decision-making for a firm, the effectiveness of Big data is conditioned by the processing of a large set of varied data, while that of Smart data depends on handling a smaller sample. It is also necessary to realize that establishing an effective Smart data policy emanates from the previous capacity of a firm to manage large and varied data quickly. Smart data is, therefore, the step after the capacity to manage massive amounts of data.

Second, the different digital policies are also part of a cycle and are interdependent. In practice, Big data is apprehended in a relatively fixed way by decision-makers. Whatever the situation, Big data is seen as the ultimate solution to data processing. Our study shows that Big data is only one proposed solution among others, such as Smart data. Thus, the context and the aims pursued condition the type of digital policy to put in place. The identification of levers of action then allows decision-makers to implement dynamic and adaptive exploitation of data (Vitari and Raguseo, 2016). In other words, it is not the tool mobilized in a digital policy that determines the aims targeted by a firm, but rather the prior identification of the strategic goals of the firm that guides the adoption of this or that tool. It is then necessary to use a data policy in accordance with the results to be achieved in order to generate added value.

6. Conclusion

Considering that the management of data has become a strategic element for the organisation, it's necessary to understand better its different traits, structures and its evolution. Our study brings us useful knowledge about the transformation of data policies with new theoretical and

practical contributions. This has revealed that small low-tech businesses are behind these many types of data management. These results highlight the characteristics of Smart data demonstrating that it is unnecessary to exploit a large volume, and wide variety of data to calculate its value. Furthermore than just the Big data model gaining its worth through literature, this new policy offers the organisation the opportunity to develop differently its digital actions by choosing to detect and exploit a smaller data volume linked to a domain that has been previously researched, studied and identified. To go on, this study has revealed that Smart data can not be a substitution for Big data. These two policies are codependent in the organisation and work in a synergistic relationship through a virtuous data exploitation loop. Future research could benefit from this study by analysing the resources and abilities that the small companies need to take in account to be able to manage better this digital transformation, therefore, the capacity to recreate and improve on the long term, the production and the quality of data.

7. Limitations and future research

We close the discussion by pointing out some limits to our study. Our quantitative method proposes a static approach to the phenomenon. A qualitative approach would complement our results by looking at the processes involved that we deliberately omitted during our study. Moreover, the results come from a questionnaire composed of items of an intuitive and subjective nature. Even if this point led to an approximation with regard to some of the answers given by the interviewees, conducting the questionnaire in this way avoided the overly precise or complex quantification of items, which may have produced a large number of non-responses.

Our research focused solely on the distribution activity of the automotive industry. However, manufacturers and original equipment manufacturers exploit data differently from the way in which dealers might. It would, therefore, be interesting to understand the digital policies of other automotive trades and to identify the determinants specific to their intra- and inter-organizational environments. This also applies to organizations that are neither small nor medium-sized enterprises and which may, for example, manage data differently due to higher endowments of resources and skills than small businesses.

It is also necessary to better understand the contexts inherent in any adoption of a digital data policy. Situations are different from one firm to another. Thus, the original 3Vs model may be obsolete and many criteria can be added according to the characteristics of firms, their strategic intentions and their resources.

References

Agarwal, R. & Dhar, V. (2014). Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. *Information Systems Research*, 25(3), 443-666. <https://doi.org/10.1287/isre.2014.0546>.

Alexander, L., Jiang, S., Murga, M. & González, M. C. (2015). Origin–Destination Trips by Purpose and Time of Day Inferred from Mobile Phone Data. *Transportation research part c: emerging technologies*, 58(B), 240-250. <https://doi.org/10.1016/j.trc.2015.02.018>.

Alles, M.G. (2015). Drivers of the Use and Facilitators and Obstacles of the Evolution of Big Data by the Audit Profession. *Accounting Horizons*, 29(2), 439-449. <https://doi.org/10.2308/acch-51067>.

ANFA (2016). Rapport d'activité 2015 [Activity Report 2015]. *Editions ANFA*, 1-2.

Ardito, L., Scuotto, V., Del Giudice, M. & Petruzzelli, A.M. (2019). "A bibliometric analysis of research on Big Data analytics for business and management". *Management Decision*, 57(8), 1993-2009. <https://doi.org/10.1108/MD-07-2018-0754>.

Baesens, B., Bapna, R., Marsden, J.R., Vanthienen, J. & Zhao, J.L. (2016). Transformational Issues of Big Data and Analytics in Networked Business. *MIS Quarterly*, 40(4), 807-818. <https://doi.org/10.25300/MISQ/2016/40:4.03>.

Bag, J., Pretorius, J.H. C., Gupta, S. & Dwivedi, Y. K. (2021). Role of institutional pressures and resources in the adoption of big data analytics powered artificial intelligence, sustainable manufacturing practices and circular economy capabilities. *Technological Forecasting and Social Change*, 163, 1-14. <https://doi.org/10.1016/j.techfore.2020.120420>.

Bag, J., Wood, L.C., Mangla, S. K. & Luthra, S. (2020). Procurement 4.0 and its implications on business process performance in a circular economy. *Resources, Conservation and Recycling*, 152, 1-15. <https://doi.org/10.1016/j.resconrec.2019.104502>.

Bag, J., Wood, L.C., Xud L., Dhamija P. & Kayikci, Y. (2020). Big data analytics as an operational excellence approach to enhance sustainable supply chain performance, *Resources, Conservation and Recycling*, 153, 1-11. <https://doi.org/10.1016/j.resconrec.2019.104559>.

Boyer, J.M. (2016). La tarification et le Big data: Quelles opportunités ? [Pricing and Big Data: What Opportunities?]. *Revue d'Economie Financière*, 4(120), 81-89. <https://doi.org/10.3917/ecofi.120.0081>.

Brown-Libur, H., Issa, H. & Lombardi, D. (2015). Behavioral Implications of Big Data's Impact on Audit Judgment and Decision Making and Future Research Directions. *Accounting Horizons*, 29(2), 451-468. <https://doi.org/10.2308/acch-51023>.

Brulhart, F. & Moncef, B. (2010). L'impact des pratiques de Supply Chain Management sur la performance de l'entreprise [The impact of Supply Chain Management Practices on the Performance of the Firm]. *Finance Contrôle Stratégie, Association FCS*, 13(1), 33-66.

Brunswick, S., Bertino, E. & Matei, S. (2015). Big Data for Open Digital Innovation – A Research Roadmap. *Big Data Research*, 2(2), 53-58. <https://doi.org/10.1016/j.bdr.2015.01.008>.

CCM Benchmark Institut (2015), « Automobile et digital : l'impact d'internet sur les pratiques d'achat », <http://www.ccmbenchmark.com/institut/blog/infographie-automobile-et-digital/>

Chen, D.Q., Preston, D.S. & Swink, M. (2015). How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management. *Journal of Management Information Systems*, 32(4), 4-39. <https://doi.org/10.1080/07421222.2015.1138364>.

Chiang, R. H., Grover, V., Liang, T.-P. & Zhang, D. (2018). Strategic value of big data and business analytics. *Journal of Management Information Systems*, 35(2), 383-387. <https://doi.org/10.1080/07421222.2018.1451950>.

Cyganek, B., Graña, M., Krawczyk, B., Kasprzak, A., Porwik, P., Walkowiak, K. & Woźniak, M. (2016). A Survey of Big Data Issues in Electronic Health Record Analysis. *Applied Artificial Intelligence*, 30(6), 497-520. <https://doi.org/10.1080/08839514.2016.1193714>.

Davenport, T.H. & Patil, D.J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*, 90(10), 70-76.

Foucart, T. (2000). Colinéarité et instabilité numérique dans le modèle linéaire [Colinearity and Numerical Instability in the Linear Model]. *RAIRO Operations Research*, 34(2), 199-212. <https://doi.org/10.1051/ro:2000111>.

Friedman, H. & Marley, A. (2015). Big Data or Small Data: That is the Question. *Workforce Solutions Review*, 11, 4-6.

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

<https://doi.org/10.1016/j.ijinfomgt.2014.10.007>.

George, G., Haas, M. & Pentland, A. (2014). Big Data and Management: From the Editors. *Academy of Management Journal*, 57(2), 321-326. <https://doi.org/10.5465/amj.2014.4002>.

Goes, P.B. (2014). Big Data and IS Research. *MIS Quarterly*, 38(3), 3-8.

Golia, N. (2013). What Big Data Means for Infrastructure Costs. *Insurance and Technology*, 5, 30-31.

Griffin, P.A. & Wright, A.M. (2015). Introduction: Commentaries on Big Data's Importance for Accounting and Auditing. *American Accounting Association*, 29(2), 377-379. <https://doi.org/10.2308/acch-51066>.

Grover, V., Chiang, R. H., Liang, T.-P. & Zhang, D. (2018). Creating Strategic Business Value from Big Data Analytics: A Research Framework. *Journal of Management Information Systems*. 35(2), 388-423. <https://doi.org/10.1080/07421222.2018.1451951>.

Günther, W.A., Rezazade Mehrizi, M.H., Huysman, M. & Feldberg, F. (2017). Debating Big Data: A Literature Review on Realizing Value from Big Data. *Journal of Strategic Information Systems*, 26(3), 191-209. <https://doi.org/10.1016/j.jsis.2017.07.003>.

Gupta M. & George J.F. (2016). Toward the development of a big data analytics capability. *Information & Management*, 53(8), p. 1049-1064. <https://doi.org/10.1016/j.im.2016.07.004>.

Holden, G. (2016). Big Data and R&D Management: A New Primer on Big Data Offers Insight into the Basics of Dealing with “Uncomfortable Data”—Data That Is Too Large or Too Unstructured to Be Accommodated by a Firm’s Existing Processes. *Research-Technology Management*, 59(5), 22-26. <https://doi.org/10.1080/08956308.2016.1208044>.

Huang, P.C. & Huang, P.S. (2015). When Big Data Gets Small. *The International Journal of Organizational Innovation*, 8(2), 100-117.

Jullien, B. (2017). Véhicules connectés à l’ère du numérique: Regards croisés sur les mutations des services de l’automobile, *Chaire ESSCA Distribution & Services Automobiles*, 1-111.

Kar, A. K. & Dwivedib, Y. K. (2020). Theory building with big data-driven research – Moving away from the “What” towards the “Why”, *International Journal of Information Management*, 54, 1-12. <https://doi.org/10.1016/j.ijpe.2016.03.014>.

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. *Meta Group Inc.*, 6, 1-3.

Li, J. & Tang, Y. (2010). CEO hubris and firm risk taking in China: The moderating role of managerial discretion. *Academy of Management Journal*, 53(1), 45–68. <https://doi.org/10.5465/amj.2010.48036912>.

Loebbecke, C. & Picot, A. (2015). Reflections on Societal and Business Model Transformation Arising from Digitization and Big Data Analytics: A Research Agenda. *Journal of Strategic Information Systems*, 24(3), 149-157. <https://doi.org/10.1016/j.jsis.2015.08.002>.

Markus, L. (2017). Datification, Organizational Strategy, and IS Research: What's the Score? *Journal of Strategic Information Systems*, 26(3), 233-241. <https://doi.org/10.1016/j.jsis.2017.08.003>.

Mcafee, A. & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*, 90(10), 60-68.

Newell, S. & Marabelli, M. (2015). Strategic Opportunities (and Challenges) of Algorithmic Decision-Making: A Call for Action on the Long-Term Societal Effects of 'Datification'. *Journal of Strategic Information Systems*, 24(1), 3-14. <https://doi.org/10.1016/j.jsis.2015.02.001>.

Pigni, F., Piccoli, G. & Watson, R. (2016). Digital Data Streams: Creating Value from the Real Time Flow of Big Data. *California Management Review*, 58(3), 5-25. <https://doi.org/10.1525/cmr.2016.58.3.5>.

PricewaterhouseCoopers (2016). Etude annuelle Connected C@r, *Strategy&*, p.1-14.

Ritala, P. (2012). Co-opetition Strategy—When is it successful? Empirical Evidence on Innovation and Market Performance. *British Journal of Management*, 23(3), 307-324. <https://doi.org/10.1111/j.1467-8551.2011.00741.x>.

Russom, P. (2011). Big Data Analytics, TDWI Best Practices Report. *The Data Warehousing Institute*, 1-6.

Schwab, A. & Zhang, Z. (2018). Editorial A New Methodological Frontier in Entrepreneurship Research: Big Data Studies. *Entrepreneurship Theory and Practice*, 43(2). <https://doi.org/10.1177/1042258718760841>.

Shamim., S., Zeng, J., Shariqb, S.M & Khan, Z. (2019). Role of big data management in enhancing big data decision-making capability and quality among Chinese firms: A dynamic capabilities view. *Information & Management*, 56, 103-135. <https://doi.org/10.1016/j.im.2018.12.003>.

Strîmbei., C. (2012). OLAP Services on Cloud Architecture. *IBIMA Journal of Software & Systems Development*, 1.

Vitari, C. & Raguseo, E. (2016). Digital Data, Dynamic Capability and Financial Performance: An Empirical Investigation in the Era of Big Data. *Systèmes d'Information et Management*, 21(3), 63-92. <https://doi.org/10.3917/sim.163.0063>.

Xerfi (2016). Le marché et la distribution automobile [The Automotive Market and Distribution]. *Editions Xerfi France*, 1-275.

Xiang, T. (2013). L'impact des coopérations en termes de ressources et d'activités sur la performance: une étude du réseau logistique en Chine [A Study of the Logistics Network in China]. *Logistique et Supply Chain Management*, 21(3), 41-56. <https://doi.org/10.1080/12507970.2013.11517024>.

Wang, G., Gunasekaran, A., Ngai, E. W.T. & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, 176, 98-110. <https://doi.org/10.1016/j.ijpe.2016.03.014>.

Zhao, J. L., Fan, S., & Hu, D. (2014). Business Challenges and Research Directions of Management Analytics in the Big Data Era. *Journal of Management Analytics*, 1(3), 169-174. <https://doi.org/10.1080/23270012.2014.968643>.