Journal of Marketing Analytics (2022) 10:187–206
https://doi.org/10.1057/s41270-021-00144-5

ORIGINAL ARTICLE

Digital footprint wrangling: are analytics used for better or worse? A concurrent mixed methods research on the commercial (ab)use of dataveillance

Bruno F. Abrantes[1] · **Klaus Grue Ostergaard**[1]

Revised: 29 June 2021 / Accepted: 19 October 2021 / Published online: 6 November 2021

Abstract
Despite the introduction in the European Union (EU) of the *General Data Protection Regulation* (GDPR) in May-18, a growing opposition is noticed against the allegedly hazardous commercial exploitation of the consumer's *digital footprints*. A "data autocracy" regime is asserted to be distorting the information to the consumer and bias the ability to take informed free choices. Consequently, companies face nowadays a great data governance challenge, i.e. to restore the consumer's positive sentiment while continuing to explore Big Data. Hence, we have conducted a multi-method enquiry with *foci* on the Danish market covering the topic of digital footprint's awareness. This was fashioned with a descriptive-exploratory research purpose, to understand the sentiment (perception) and the behavior (action) of the data-owners and data-brokers surrounding the dataveillance over personal lives. Results confirmed a generalized inability to minimize risks of data misuse. A new insight for Marketeers is the willingness of some respondents to pay for security services and thus safeguard their privacy (disbelieved of the regulatory compliance system). A perception-behavior gap is noticed on the distress with the exposure of personal information and the paradoxical low self-defensiveness. Likewise, data-brokers admit the deliberate use (perception) of imitation strategies (e.g. hyper-targeting; algorithmic refinement; and, predictive modeling) for maintaining a competitive parity with other firms, which contrasts with the institutional isomorphism (inaction) to interrupt them. Given the relevance of this discussion (and our conclusions) for policy-making, managers (inclusively marketing professionals) and citizens, is recommended the deepening of this research line in the region, especially in Nordic countries.

Keywords Big data · Consumer's sentiment · Denmark · Digital footprints · GDPR

JEL Classification M15 · M31 · M38

Introduction

Background and initial problematization

Indeed, the desire of the organizations to deepen the capacity to accommodate digital information was accompanied by an exponential upward tendency of new data colliding in the 1980s with the mounting constraints regarding data warehousing infrastructures (Hilbert and López 2011). Hence, Big Data (BD) emerged in the 1990s as the *data boom*, defining the amount of data greater than the capacity of what a computer could process or store, nor their analytics performed by standard software, regardless of its configuration (Cox and Ellsworth 1997; Mayer-Schönberger and Cukier 2013).

Beyond the benefit of solving data-related constraints, Big Data opened a myriad of opportunities for information building (Abrantes 2020). With the rapid development of a triad of (*advanced; exploratory* and, *discovery*) *Big Data Analytics* techniques, a broad instrumentalization potential has emerged and its commercial use came naturally (Russom 2011). In the first decade of the twenty-first century, the BD resembled the "holy grail" of competitive advantage and many firms have re-strategized their position in business (Lambrecht and Tucker 2015).

✉ Bruno F. Abrantes
btfa@niels.brock.dk

[1] Niels Brock Copenhagen Business College (NBCBC), Bispertorvet, 1-3, 1167 Copenhagen, Denmark

BD allowed the creation of associative patterns between units of information, such as, meaningful data systems. This triggered the fashioning of modern business *analytics strategies* for maximizing market intelligence. Firstly, companies were able to remedy their inner vulnerabilities (enterprise-wide) and simultaneously leverage their system's operational efficiency. Secondly, it democratized the use of data-driven insights for the gain of competitive parities or advantages (Boyd and Crawford 2012). An example of the utility of BD analytics is nowadays the prevention of defects on vehicles, since the handling of data from sensors measuring multiple parameters of functionality (e.g. temperature, vibrations, oil pressure, or fuel consumption) bypasses eventual breakdowns before they actually occur.

In this context, BD has expanded horizons of analyzability but also transforming the analysis function itself, which shifted their predictions from the rigor of sampling techniques to focus on the factum, as Big Data dealt with a sheer volume, almost entirely unstructured and heterogeneous, with the scale compensating data inaccuracies, unraveling patterns, forecasting scenarios and eliminating human bias and error (Baruh and Popescu 2017; Gandomi and Haider 2015).

Unsurprisingly, the virtues of BD spawned a wave of euphoria labeled as *dataism,* intrinsically assimilating the level of trust assigned to businesses and public institutions pertaining to the use of private data. Within the myriad of units of petabytes exchangeable every day, it is also the personal data deriving from the individual's internet navigation through online channels (i.e. *digital footprints*) (Muhammad et al. 2018).

In recent years, an emerging "anti-dataist" discourse has emerged, concerning the rising *datafication* processes, targeting, particularly, the overuse and misuse of personal data and digital footprints. Such perspective emphasized the increasing manipulation of Big Data, such as life events or discrete social interactions, as valuable and monetizable insights for the firms, and potentially hazardous and noxious for the individual, here identified as the personal *data-owner* (Van Dijck 2014; Gandomi and Haider 2015).

In this regard, Zuboff (2019) termed the prior controversy (dataism) as a totalitarian pattern of a surveillance capitalism (*dataveillance*) imprisoning citizens and subordinating those to behavioral modifications. Accordingly, the ability to control behaviors and predict events is a valuable asset allowing the tradeable of separate units of data, equivalent to human stocks. *Data-brokers,* as the entities responsible for the direct and/or indirect manipulation of those human stocks are perceived as hazardous entities with a questionable moral conduct. This occurs, not solely on the grounds that organizations are able with BD analytics to anticipate future actions; but foremost, because of the asserted anti-democratic power to alter what one is able to think, see or feel. Market wise is a deceptive way of modifying the patterns of choice and consumption (Ahn and Lee 2020; Dekimpe 2020; Mayer-Schönberger and Cukier 2013; Leinweber 2007).

This constitutes the initial problematization for pursuing such a research avenue, i.e. to understand how individuals and firms perceive (sense and feel) the access and utilization of digital footprints and the consequent behavior adopted by both parts. A literature review will expose an alleged dispute of individuals and firms and the rationale for such conflict of interests.

Research gap and aims/objectives

As an emergent field, BD earned a growing audience of academics and rise on publications in recent years, with a significant coverage of digital footprints, namely as to sentiment analysis. Nevertheless, the studies are few on the Nordics, especially on the Danish market, cross-observing *rationem* (reason) and (*motus*) emotions, regarding the utilization of personal data's perceptions (digital footprint's awareness and sentiment) and individual behavior (action and self-defensiveness). Therefore, the uniqueness of our research resides on the joint investigation of these two underlying components, i.e. data-owners and brokers. Moreover, the firm's perspective (perception and behavior) is a fairly untapped research angle in Denmark. This would justify per se the choice of the marketplace of observation; however, the country is a highly digitalized society which constitutes an ideal scenario for running such test.

Hence, our purpose is to grasp whether the individuals acknowledge and/or accept the utilization of their personal data by third parties. Likewise, our enquiry mirrors the same components on the data-broker's side. Thus, the general aim (A) of this research is to understand and explore perceptions and behavior of individuals (data-owners) and firms (data-brokers) in relation to their mutual data exchange. Therefore, a set of objectives is derived from the above:

Objective 1 (O1—Data-owners' perceptions): To comprehend the consumer's general perception of data-broker's conduct processing their personal data.

Objective 1.1 (O1.1—Data-owners' perception versus action): To identify behavioral consistency or deviation in attitudes.

Objective 2 (O2—Data-brokers' perceptions): To grasp the data-brokers' knowledge about the consumers' perceptions and sentiment.

Objective 2.1 (O2.1—Data-brokers' perception versus actions): To determine whether firms adapt the utiliza-

tion of personal data based on consumer's degree of avowed (un)satisfaction/(un)willingness.

The figure represents the association of aims, objectives and the formulation of research questions. A general aim is divided into four specific objectives, of which two general objectives (O1, and O2) and two specific objectives (O1.1 and O2.1.) deriving from the first ones. The general objectives cover the perceptions of both data-owners (O1) and data-brokers (O2) while the specific objectives address the gap between perceptions and in/actions. A multi-case research with a comparative design utilizes embedded units for testing the above framework. A one-phase mixed methods study instrumentalizes two methods with a concurrent mode of collection, as Danish firms and final consumers' perceptions and actions are scrutinized through internet-mediated interviews and questionnaires with the researchers' assistance (Fig. 1).

Accordingly, the general structure of this manuscript proceeds with a theoretical revision of the literature on Big Data Analytics on spatial BD emphasizing the digital footprint's utilization of personal data focused on the knowledge and behavioral- related aspects of awareness and sentiment. The article proceeds with a description of the research design, the typology of case research, the demographic profiling of the participating firms and sampled units of analysis, and with the clarification of the rationale for the choice of methods. Subsequently, the research continues with the application of methodological procedures to data analysis, extrapolation of results from data manipulation and a discussion, as to the findings deriving from the data outputs. Finally, the conclusions of the study are presented in connection to the aim and objectives, complemented by the managerial implications and the opened avenues for further exploring this topic and subsequent knowledge-building.

Literature review

Big Data is a phenomenon of increasing ubiquity and henceforth, literature associated with the phenomenon is becoming inherently copious (De Mauro et al. 2015). The vastness of the theme led us to immerse into a theoretical revision centered on the initial problem addressed in the previous section, i.e. the sentiment of digital footprint's (ab)use. However, we start by conveying first a consensual definition of BD (De Mauro et al. 2015, p. 103):

> Big Data represents the Information assets characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value.

This definition, covers most components addressed in others and conveys four discernible components:

1. Information: the fuel of BD, requires organizing/cross-referencing to generate insights.

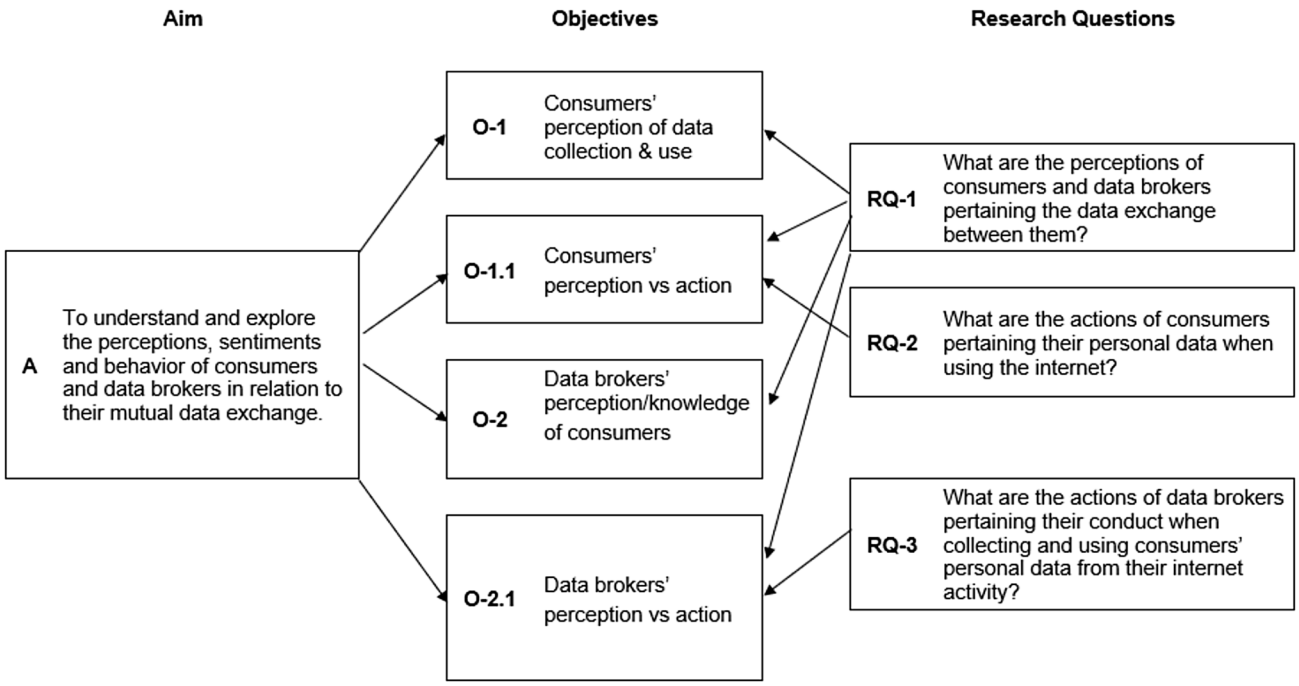

Fig. 1 Research framework. *Source* Own elaboration

2. Technology: The exponential increase of data with the proliferation of Internet of Things (IoT) requires an equally exponential increase in data storage and data processing capabilities.
3. Methods: The bottleneck of BD lies presently on the actual analysis.
4. Impact: Actual and potential (impact) across multiple industry areas and science.

The forth component (impact) refers to the three-dimensional perspective of BD (termed as the "3 V's"), a combination of societal advancements and hastening of BD analytics (i.e. volume, velocity and variety), which opened new horizons for a data-driven economy (with a rising market value), labeled as the Second Economy (Lerman 2013; Manovich 2011; Jacobs 2009).

Therefore, BD is acknowledged as a key economic asset for the future, and an instrument for competitive advantage-gains; however, with challenges for the firm's ability to acquire new skills and incorporate new technologies; especially as to the scalability (volume) to accommodate drastic changes in magnitude one can manipulate, on a fast pace (*velocity*) and on a broad span of informational assets (*variety*) (Cavanillas et al. 2016; Kitchin and McArdle 2016).

These datasets with massive quantities of data (large volume) generated in a continuous way, on real-time basis (velocity) and, accommodating variety (structured/unstructured and semi-structured) are also exhaustive records of the entire population, rather than samples, capturing entire systems of behavior (Mayer-Schonberger and Cukier 2013). Furthermore, these datasets provide a tight and fine-grained instrumentalization (*resolution*) and uniquely indexical (*identification*) of their significance (Dodge and Kitchin 2005); with a strong *relationality* as to the conjoining with other datasets; *extensionality*, allowing the flexible incorporation of new fields and expansion in breadth of coverage (Marz and Warren 2012).

Data-brokers are unsurprisingly called nowadays the "wizards of mining" (Marr 2017). It is undeniable the dependence of BD, as a key economic asset (though value-free) to BD analytics. The latter is the key activity responsible for the effective value creation, yielding insightful depth of information extracted from units of petabytes. On the next section, are here introduced the main phenomena surrounding the problematization of the commercial exploitation of personal data and contextualized into EU's legal framework.

Digital footprints: wrangling and challenges

To arbitrate the European Big Data Ecosystem (EBDE) a new legal framework from the European Parliament and Council's side was issued on May 25th May 2018, through regulation for the protection of unrestricted movement of such data. This legal act was designated as the *General Data Protection Regulation* (GDPR)—*Regulation (EU) 2016/679* (EU 2020; Eur-Lex 2016; Barlow, 1996), including, among others, the instruction on *data portability* (relationality); *data breaching and data sharing, consent and transparency* (resolution and indexicality*)* (De Hert et al. 2017; Abrantes and Venkataraman, in press). Therein, *article 5* established seven key-principles of data protection policy: (a) lawfulness, fairness and transparency; (b) purpose limitation; (c) data minimization; (d) accuracy; (e) storage limitation; (f) integrity and confidentiality; and, the accountability principle. Despite, the main similarities to the earlier legal regime, i.e. the Data Protection Act 1998 (*1998 Act*), the GDPR neglected a principle for the protection of individual rights as to the international transference of personal data (ICO 2018).

Such an omission in the previous regulation is attempted to be disentangled as to personal data protection, as the principles above, are extended on November 21st on the *Official Journal of the European Union* (L 295/39) with the *Regulation (EU) 2018/1725* which came to force on the 28th October 2018 emphasizing that,

> The protection of natural persons in relation to the processing of personal data is a fundamental right. Article 8(1) of the Charter of Fundamental Rights of the European Union (the 'Charter') and Article 16(1) of the Treaty on the Functioning of the European Union (TFEU) provide that everyone has the right to the protection of personal data concerning him or her. (p. 39)

Regardless of such legal framework, a criticism towards the gap between law making and supervision is seemingly growing in EU, especially as to the commercial usage of personal data by corporations (as presented at the "Introduction" section). Such friction seems to be triggered by a confrontational scenario, fueled by the firm's practices of *hyper-targeting* of consumers, in conflict with privacy rights (including individual ones) observed in the acts above (Crawford and Schultz 2014).

Furthermore, the GDPR touted in *article 17* the "right of erasure" or the "right to be forgotten (RTBF)" described below. Likewise*, article 37* defined the creation of a mandatory role, the *Data Protection Officer (DPO)* for all organizations acting as *data-brokers*, which core activities involve regular and systematic monitoring of personal or sensitive data on a large scale (ICO, 2018; EDPS, online). Here, the notion of data-broker covers both the concepts of *data suppliers* and *technology providers.* The first *(data suppliers)* referring to the "Person or organization [Large and small and medium-sized enterprises (SME)] that create, collect, aggregate, and transform data from both public and private sources" (Cavanillas et al. 2016, p. 52). The second (technology providers) as "typically organizations (Large and SME) as providers of tools, platforms, services, and know-how for data management." (Cavanillas et al. 2016, p. 52).

Such *hyper-targeting* (or *micro-targeting)* gives attention to an exhaustive utilization of detailed customer-data and marketing automation tools, based on predictive modeling, statistical tools, and algorithms to target and deliver highly personalized messages across a large number of digital channels (Semerádová and Weinlich 2019). This represented another underlying assumption about Big Data, which is, on its own BD is value-free unless algorithmically transformed into "actionable insights" which in turn can be applied to solving problems in a wide variety of applications by identifying patterns imperceptible to human logic and thereby predict future conditions (Baruh and Popescu 2017).

However, the controversy towards digital footprints, reveals an exploitation of data with both commercial and non-commercial purposes (Muhammad et al. 2018). We recall here one of the most striking examples of the latter, the 2015s *Cambridge Analytica* data (breach) scandal on the North-American presidential campaign, with the harvesting of social media information without personal consent (Demant 2019; González 2017). Nonetheless, some scholars raise doubts as to the absolute effectiveness of BD, as some studies of BD's advertising performance exhibit contrary results. Here, Semerádová and Weinlich's (2019) research of over 840 Facebook ads, concluded in fact a limited effect and counterproductive results in terms of user reactiveness. This suggests a negative correlation of overexposure to hyper-targeting and future purchasing causality, as the first may lead to negative attitudes toward the advertiser. Yet, a negative sentiment finds real ground implications on privacy loss, self-determination and price discrimination. Hence, some scholars have postulated the importance of "stopping the unjustified accumulation and commercialization of personal data" and furthermore advocating a higher responsibility of consumer protection, which ought to be vested in the legislator hands to immediately regulate and execute the compliance of practitioners to it (Lecuona and Villalobos-Quesada 2018, p. 291). These author's assertion comes in a current scenario where further concerns were raised towards the public sectors' equivalent exploitation of BD.

Yet, regardless of the ability of the data-brokers to determine future outputs of on consumers' actions, the instrumentalization of human behavior is nowadays unavoidably criticized as a hazardous use of digital technology towards mind-reading, as an abusive anticipation of human behavior and personality profiling (González 2017). Here, the author stresses furthermore the peril of a malevolent instrumentalization of artificial intelligence (AI) for the manipulation of the masses. Conversely, consumers are further perceived furthermore as being unprepared to understand or evaluate the ethical implications and the knock-on effects on their lives of the usage of their digital footprints (Zwitter 2014). For instance, personal life actions as "feeds" and ''likes'' used by Twitter and Facebook's analysis of sentiment, group manipulation and micro-targeting, working as monetary units of value for marketing companies. This instrumentalization of personal data has been associated with a furtive molding of consumer's thoughts, and their intentional redefinition of behaviors (causality reasoning). Therefore, some scholars have alluded to a form of digital captivity or servitude, as a current state of Big Brother's social regime, of post-modern digital voyeurism and surveillance capitalism of (quasi) totalitarianism (Zuboff 2019; Diamond 2019; Brayne 2017; Xue et al. 2016).

Whether they be commercial, symbolic or political issues relating to Big Data analytics within the EU, the issues themselves become more pertinent due to the GDPR. Jones (1991, p. 367) defined business ethics as an attribute of a decision that is both legal and morally acceptable, and unethical "is either illegal or morally unacceptable to the larger community". The EU's GDPR, as a legal framework attempting to promote digital democracy, proclaimed in article 17 the citizen's *Right to be Forgotten* (RTBF), which then spurred a worldwide debate as to the right of the individual to privacy (Xue et al. 2016). The RTBF includes the supreme right to own digital anonymity. Thus, the data-broker is obliged upon requested to delete all personal data without undue delay. This latter agent retains in this case no overriding legitimacy to further process of data, being such action considered an illegal act.

Nevertheless, companies seem to comply only partially with the GDPR. For instance, RTBF requesters of delisting (e.g. from a company's newsletter) face the unprotected assault of data reutilization by third entities (*Streisand effect*). Whenever consentment is once provided to them, the annulment of the decision is in practice irreversible. Even under the eye of the compliance authority (i.e. the European Data Protection Supervisor) the data-sharing chains are hardly undone and the person is unable to fully exercise the right to anonymity, whether to hide, to remove, or to censor a piece of information, (Xue et al. 2016; Thatcher 2014).

Moreover, firms use analytics to yield new data. The original inputs from individuals are also reconditioned into new self-generated data bundles deriving from the interpretation of patterns of behavior and further monetizable predictions (*data fumes*). In this way, natural data obtained from online "base behaviors" face again the threat of undue exposure, as their reutilization through data fumes correspond to an exhaustive extraction procedure of third-party vendors with inherent dangers of building biased contrived datasets, with a likely distance from the real demonstrations of behavior (Thatcher 2014).

Consequently, the GDPR created a mixture of disappoint and skepticism among EU citizens, due to the EU's inability to ensure a full compliance of principles and dispositions, namely written on the articles 17 and 37. Furthermore, as the article 17 forces the controlee to take action on his/her hands

and report abusive practices to the supervisory authority to act by complaint, the regulatory system is questioned as to the RTBF's effectiveness. The inaction of the consumer is penalized and stimulated the data-broker's prevarication and non-compliance. This constitutes one of the objectives of the study *(O.1.1.consumer's perception vs. action)* to comprehend the dissonance between digital footprint's sentiment and individual actions taken.

Digital footprint's awareness

Consumer's background

Through data-driven marketing, consumers began to experience empowerment through optimal satisfaction of personal preferences (André et-al. 2018). Big Data provides insights into consumer preferences, which in turn enabled the industry to target individual customers with personalized options or recommendations better fit to satisfy them. Consumers enjoy nowadays an eased shopping experience, time-saver and focused on preferential choices, which translates into convenience and lower resource consumption (Chen et al. 2012). Furthermore, consumers benefit from an abundance of internet-based free of charge services, such as communication through social media, access to endless sources of information and entertainment enjoyed by most people.

On the other hand, data-driven marketing backfires through pigeonholing consumers into behaviors, often referred to as *consumer welfare depreciation* (André et al. 2018). Data-driven marketing naturally focusses on behavioral anticipation, and not necessarily on higher-order psychological processes (e.g. emotions, moral judgements, preferences or "meta-preferences") being customers *aspirational preferences* differing from their actual behavioral (determined) *preferences* (André et al. 2018).

Van Dijk (2014) argues also that a loss of privacy and furthermore the normalization of privacy loss, is exemplified by the common practices of the consumers (e.g. accepting cookies on websites) without actually measuring the associated privacy challenge and assessing the privacy policy of a website, and thereby giving unrestricted consent to their data. Baruh and Popescu (2017) argue that this is a matter of structural failure instead of a lack of individual skills. These authors challenge then the notion of *privacy* as being incorrectly constructed as an exclusive individual concern and responsibility when it is in fact a collective value with a collective social dimension. Regulatory efforts should therefore consider privacy in the digital domain not by exclusive self-management by the individual but by government's more effective legislation. Furthermore, Big Data is enabling not merely *Price Steering* but *Price Discrimination*. *Price Steering* as to the personalized content in e-commerce, as two consumers using the same search string for the same product receive different product results, or also, the same results presented in a different order, according to the algorithm's prediction of the relative affluence of the consumer. This is a way of nudging the consumers towards products of higher value, and with a higher price. Subsequently, the *Price Discrimination* entails the differential offer of prices to potential consumers for the same product. The more affluent consumer is shown a higher price for the same product than the less affluent consumer (Hannak et al. 2014). Here, a significant threat is the *First-Degree Price Discrimination* aiming at micro-target the consumer's maximum price or reservation price. This refers to the mapping of the maximum limit of willingness of the consumer to pay for a certain product.

Typically, this is theoretically possible only in a monopolist market structure, but Steinberg (2020) asserts that advancements in Big Data identifying individual consumer's reservation prices make it possible even in a perfect competition scenario. Yet, Mayer-Schönberger and Cukier (2013) assert that the analytics' exacerbation toward price controlling or fixing is more likely within central planning systems.

Data-broker's background

Approximately ¾ of the surveyed companies argue Big Data represents an opportunity, as analytic algorithmization yield a variety of benefits ranging from the simple understanding of customer behavior until a better segmentation and targeting for tapping into new market opportunities (Russom 2011). Thus, harnessing big data effectively may represent additionally a competitive edge (Mayer-Schönberger and Cukier 2013). The International Data Corporation (IDC) asserts that the current digital transformation has pushed the IT industry towards a *third platform technologies* phase, which syndicates (and intensifies) mobile and cloud computing; social media, the internet of things (IoT) and BD analytics. This anticipated the "digisense" era of abundant interrelatedness of sensors, controllers, big data and data science (Gartner 2019).

Consequently, the market size of this technologies is expected to reach the USD 6 trillion USD mark by 2022 (IDC 2018). For instance, Walmart which is the large retailer in USA is also a pioneer on BD analytics on its industry. Walmart collects already 2.5 petabytes (2,500,000 gigabytes) of data per hour and applies real-time analysis on internal sources (e.g. product turnover, customer transactions, financial data and customer traffic) and external sources (e.g. social media comments, mobile phone data, e-mails, website clicks, weather and temperature). Such inputs allow the targeting of recommendations, in-store navigation, improvements on merchandise display, and the optimization of the supply chain processes, as to the price

negotiation, and maximization of profit pools (Benjelloun et al. 2015; Marr 2017).

The traction gained by Big Data on the public administration favored also a variety of new applications ranging from education to health, increasing citizen's engagement in public affairs, prevention fraud/crime, improving national security, and supporting other forms of well-being (Kim et al. 2014). For instance, the New York City (NYC) fire department developed a whole new fire-prevention strategy gathering 900.000 buildings' typologies on a single dataset correlated with tax information, ambulance visits, local crime rates, rodent complaints, and records of fires in buildings, to establish fire risk predictors, anticipating incidents and optimizing the daily inspection work (Mayer-Schönberger and Cukier 2013). However, prior literature denotes a mix of externalities: strengths, weaknesses; opportunities and threats (SWOT) associated with both the Big Data analytics by both public and private organizations. This raises ethical issues as to its adoption and uncovering the debilities of the GDPR, namely the insufficiency of compliance supervision (Xue et al. 2016; De Mauro et al. 2015; Thatcher 2014).

Methodology

The research framework diagrammatically represented in "Research gap and aims/objectives" section aggregates the tenets of the theoretical testing here followed, as the researcher team intended to gain an insight as to the extent of conformity between sentiment and behavior towards digital footprints. Moreover, it aims to unveil whether consumers are willing or not to waive own rights and their efforts to ensure that these are effectively met.

We have adopted a post-positivistic stance, using a one-phase mix-methods with a concurrent application. Such research decision derived from a seminal reflection on a triad of considerations: the aims/objectives; the assumptions of the research team; and, the ethical commitment to this empirical testing. In this context, a multiple case research with a comparative design focused on two large technological Danish firms, with the incumbents classified as Firm 1 (F1) and Firm 2 (F2) for anonymity purposes. From those, quants and qual data was collected on and about these firms, respectively from the senior managers representing the data-brokers' side, and from the final consumers corresponding to the data-owner's side.

Quantitative data (quants) obtained from 137 eligible questionnaires (Q_n), applied to the data-owners' side, targeted potential respondents through a social media platform, using personalized messages. As to the latter, the response rate achieved a high mark (0.9648), as observed with 5 partial questionnaires excluded covering between > 50% and < 100% of the questions. No cases observed of break-off records (< 50% of response).

Table 1 Case-firms (F_n)

Factor	F_1	F_2
Establishment (year)	1904	2014
Employees*	3,0	2750
Revenue (DKK)**	1,9	11,67

Source Own elaboration

*Expressed in Thousands of units (K)—Includes permanent/temporary employees (year/$\bar{x}$)

**Billions *Danish Kroner* (DKK)

Table 2 Participants (P_n) and Respondents (R_n)

Factor	(P_n)		$R(_n)$
	F_1*(P_1)	F_2*(P_2)	–**
Age	–	–	46–65 (*Mo*)
Education*	7	6	–
Seniority in firm	19	10	–
Job position	VP[a]	HC[b]	–

Source Own elaboration

*Levels in accordance to the European Qualification Framework (EQF)

[a]Vice-President, Business and Portfolio Planning

[b]Head of Controlling & Real Estate

The quants is further discussed in "Data-owners" section. *Data-Owners* and represented as to an econometric formulation of the model as to the extent of a relation between two parts: perceptions and behavior, given respectively by opinion variables and action variables. Thus, the development of an appropriate method for the measurement of these relations, is set on the assumption that the (relationships of the) individuals' endeavors to ensure an optimal degree of privacy (or behavior) constitute the *y* type of variables (action), and the opinion hold about the privacy of their digital footprints (perception) the *x* variables.

Qual units of analysis (UA) applied to each case-firm represent the data-brokers' side, and multiple quantitative ones obtained from their customer-base, configured the use of identical analytical procedure between cases, assuming naturally an iterative logic with embedded UAs. In total, 2UAs were collected from the case-firms using open-end interviews (I_n). The profiling of both case-firms and the participants/respondents is presented in Tables 1 and 2. Moreover, a pre-testing was conducted on Feb-20. The first method was applied, as to the collection momentum, as a one-off vis-à-vis interview conducted respectively on March 6th (F_1) and March 10th (F_2). The second method was internet-mediated and self-administrated during the same timeframe,

a 20 days' time around was given from counting from the invite for further enquiries and subsequent delivery.

The selection of the cases followed a (non-probabilistic) purposive and snowball logic, including as to the accessibility to the senior manager on case-firm F2. Conversely,

The participants P_1 (of F_1) and P_2 (F_2) adjusted to the same age group of the modal class of the respondents, which contained 93 respondents. The mean of seniority ($\bar{x} = 14{,}5$) of the participants corroborated the formal education background ($\bar{x} = 6{,}5$) supports their purposive selection as potentially insightful participants.

As to the design of the collection methods, the one applicable to senior managers, as data-brokers, used prompting as technique through an interview guide containing 10 questions, and with no intertwining of any probing. To the analysis of the respondents, as data-owners, opinion and behavioral variables were used to comprehend both sentiment and actions, through 14 investigative questions (IQ) being safeguarded the seminal tenets of anonymity of the individual, confidentiality of answers and the restriction of access to data records to third parties. Figures of enquiry above exclude in both methods the profiling/demographic questions above summarized.

Both methods were fashioned in Danish language with subsequent retroversion of results to English language.

The signifiers of the data-brokers interviews (Exhibit 1) were audio-recorded, transcribed and converted into to English for the analysis of manifest content, as the content verbalized by the informants, on the light of thematic analysis (TA) while data-owners answers were manipulated using statistics tools for the generalization of their results. Both are, furthermore exposed in "Data analysis and discussion" section. *Data analysis and findings.*

Data analysis and discussion

Data-owners

According to the purpose of this study *(O1; O1.1; O2; O.2.1)* and testable propositions *(RQ1; RQ2; RQ3)* exhibited at the research framework (*1. Introduction*), the data-owner's questionnaire was divided into three categories of variables, as exhibited below (Table 3):

Type 1 or *attribute variables (var001x: navigation)*, Type 2 or *opinion variables (var002x: online activity)*, and Type 3 or *behavioral variables (var003x: exposure-willingness)*.

Hence, a theoretical model with the linear function *(f)* is chosen as,

BAGGRUNDSINFORMATION

1. Køn

- Kvinde
- Mand
- Andet
- Ønsker ikke at oplyse

2. Alder

- 15 år eller yngre
- 16 - 25 år
- 26 - 45 år
- 46 - 65 år
- Over 65 år

5. Hvilke type(r) aktiviteter bruger du internettet til?

	Dagligt	Ugentligt	Månedligt	Halv-årligt	Aldrig/næsten aldrig
Shopping	☐	☐	☐	☐	☐
Spil/gaming	☐	☐	☐	☐	☐
Gambling	☐	☐	☐	☐	☐
Motion/fritid	☐	☐	☐	☐	☐
E-mail	☐	☐	☐	☐	☐
Chat/Messenger	☐	☐	☐	☐	☐
Dating	☐	☐	☐	☐	☐
Nyheder/Sport	☐	☐	☐	☐	☐
Uddannelse	☐	☐	☐	☐	☐
Arbejde	☐	☐	☐	☐	☐
Netbank	☐	☐	☐	☐	☐
Sociale medier (Facebook, Twitter, LinkedIn etc)	☐	☐	☐	☐	☐
Andet	☐	☐	☐	☐	☐

Exhibit 1 Excerpt of method 2 (quants—questionnaire). *Source* Own elaboration

Table 3 Data-owner's categories and variables

Category		Variable		O → RQs
ID	Description	ID	Description	
Attributes (1)	Profiling	*Var001*	*Time expenditure*	–
Opinions (2)	Awareness	*Var002-1*	*Convenience-Exposure*	*O1 ~ RQ1*
	Sentiment	*Var002-2*	*Exposure-willingness*	*O.1.1 ~ (RQ1/RQ2)*
Behaviors (3)	Activity	*Var003-1*	*Navigation track*	*O2 ~ RQ1*
	Defensiveness	*Var003-2*	*Self-protection*	*O2.1 ~ RQ3*

Source Own elaboration

$$f = \left(X_{1,}X_{2,\ldots,}X_{k,}\right) = \left(X_{1,}X_{2,\ldots,}X_{k,}\right) \quad (1)$$

With an empirically testable form, the specification of the stochastic structure of the variables of the linear regression is explained by:

$$y = f\left(X_{1,}X_{2,}\beta_1\beta_2 + \cdots + X_k\beta_{k,}\right) + \varepsilon \quad (2)$$

In which, the approach to determine the linear model consists of,

$$y = f\left(X_{\text{data-owners opinions,}}\beta_{\text{data-owners actions}}\right) + \varepsilon \quad (3)$$

$$\text{Activity}_{(y1)} = \alpha_0 + \beta_1\text{awareness} + \beta_2\text{sentiment} + \varepsilon \quad (4)$$

$$\text{Defensiveness}_{(y2)} = \alpha_0 + \beta_1\text{awareness} + \beta_2\text{sentiment} + \varepsilon \quad (5)$$

Here, the statistical model accounts a residual ($\varepsilon > 0$) accounting for the error of judgement of data-owners. The partial derivatives of y are considered independent ($i = 1, 2, 3\ldots$) with respect to each of the parameters $\beta_{1,}\beta_{2,}\ldots\beta_{k,}$.

The attribute variables referred to the user's profile as to the attention placed on internet-mediated activities *(var001-time expenditure*). The opinion variables express both awareness and sentiment towards navigation. The first addresses (awareness) depicting its utility (*var002-1: convenience of exposure*) and the latter (sentiment) the perspective of the individual towards own data commercial exploitation *(var002-2 exposure-willingness)*, here including data fumes and Streisand effect. The behavioral variables emphasize the actions taken online, as the patterns of navigation *(var003-1: Navigation track)* and proactive defensiveness of own rights (*var003-2:-self-protection*).

These five variables allowed a dyad of data triangulation procedures within primary data. First, a crosschecking of demographics (e.g. gender; age; or, education) with the answers to IQs. Furthermore, a triangulation of the variables Type 2 with the variables Type 3, for understanding its conformity, between awareness-activity (conformity model 1); awareness-defensiveness (conformity model 2); sentiment-activity (conformity model 3); and, sentiment-defensiveness (conformity model 4); and therefore answer the RQ3 as to association of perceptions/opinions with behaviors.

Opinion variables were treated by four investigative questions (IQ*# 9;* IQ*# 10;IQ#11;* IQ*# 14)*. One on awareness ((IQ*# 9)*, emphasizing the utility or convenience versus nuisance (convenience-exposure) ("What is your opinion of following statement: "The collection of my personal data, in general has a positive impact on my convenience when using the internet"?). The remaining is on sentiment, emphasizing confidence, sentiment and openness to further paid services. The IQ#10 focused on the confidence of utilization (exposure willingness) using question ("How would you describe your trust on having your personal data processed legally using your own activities on the internet?"). The prior is articulated on sentiment by IQ#11 ("What is your opinion on, having your data being for targeted advertising and personalization of search results?"). But also with openness to paid services with IQ#14 ("Would you be willing to pay a small amount, in order to avoid your personal data to be collected and sold?").

Data-owner's responses to IQ#9, denoted a significant level of disagreement as to the convenience of utilization of their internet navigation patterns, with a relative frequency of 0.51 (disagreement) against 0.38 (neutral) and 0.11 (agreement). The "neutral" scale classification describes the stimulus of "Both Positive and Negative"; "Neither Positive nor Negative", as well as, "I don't know". As to the sentiment (var002-2), the question IQ#10 about the level of confidence with a legal processing of data, responses had an identical result with 0.51 of disagreement; 0.24 (neutral) and 0.25 (positive). In addition, IQ1#11 revealed an accentuation of the "annoyed" to a relative frequency of 0.7, while 0.22 were neutral and 0.09 had a positive opinion about their personal data being collected and used for marketing purposes.

Moreover, a confirmatory question IQ#13 ("How do you consider targeted advertising on the internet?") tested the internal consistency of the results on IQ#11. The mean difference between the two groups of results was of 0.049 and the Cronbach alpha of 0.83. Answers to IQ#13 confirmed similar results, with respondents assuming to be 0.65 (annoyed), 0.25 (neutral) and 0.1 (positive). Furthermore, enquired about the future, the responses to the question

IQ#15 ("how do you think your personal data will be processed in the future?") the differences of opinions seem more meager, with 0.41 of respondents claiming to be "less secure", 0.26 "neutral" and 0.33 "more secure". Again, asked about the future, as to the likelihood of paying for securing the non-utilization of their data, results exhibited clear negative opinion as 0.22 answered "Yes", 0.2 "I don't know", and 0.58 answered "No".

The interpretation of the results from interval variables of the respondents' answers to IQs instrumentalized the correlation coefficients, as to Evans (1996) levels of significance. The latter varying from a very strong negative (− 1 and − 0.8) to very weak negative (− 0.19 and − 0.01) correlation; and from very weak positive (0 to 0.19), moderate (0.4 to 0.59), to very strong positive (0.80 to 1) one. Here, the results on opinion variables as to convenience *(var002-1)* revealed a respondent's acquiescence bias as he results (0.51) illustrate neither a positive/negative sentiment on sharing of personal data, while 31% neither agree/disagree at all. Conversely, 18% demonstrated a clear positive judgement of agreement. As to the sentiment variable *(var002-2),* the tabulation of demographic data with sentiment revealed no significant correlation, as to gender-sentiment ($r = 0.09$) and age-sentiment ($r = -0.05$). Yet, regardless of their profile, either more than 70% of the respondents consider "annoying" or "very annoying" their commercial utilization, when enquired about (IQ #11) "What is your opinion on your personal data being collected and used on targeted advertising and personalization of search results?". However, answers to IQ #14 ("Would you be willing to pay a small amount, in order to avoid your personal data to be collected and sold?"). The majority of the respondents (0.58) are unwilling to pay for internet services to avoid personal data collection and its commercial exploitation. Despite the weak correlation coefficient per gender, women revealed though a more openness to pay for internet services ($gender_{(female)} = 2 \Leftrightarrow r_{(2)} = = 0.13$) against ($gender_{(male)} = 1 \Leftrightarrow r_{(1)} = = 0.11$) but to comprehend its statistical significance the sample required to be added representative features to the target population.

The behavioral variables emphasized as to the typology of activity, as chatting, dating, education, email utilization, gaming, gambling, net banking, news, shopping, social media usage or other purpose (Exhibit 2).

As to the self-protection, respondents indicated which protective measures they have applied and how often split between the following options (Table 4).

The majority acknowledged that they never (or almost never) have applied such protection measures. The least

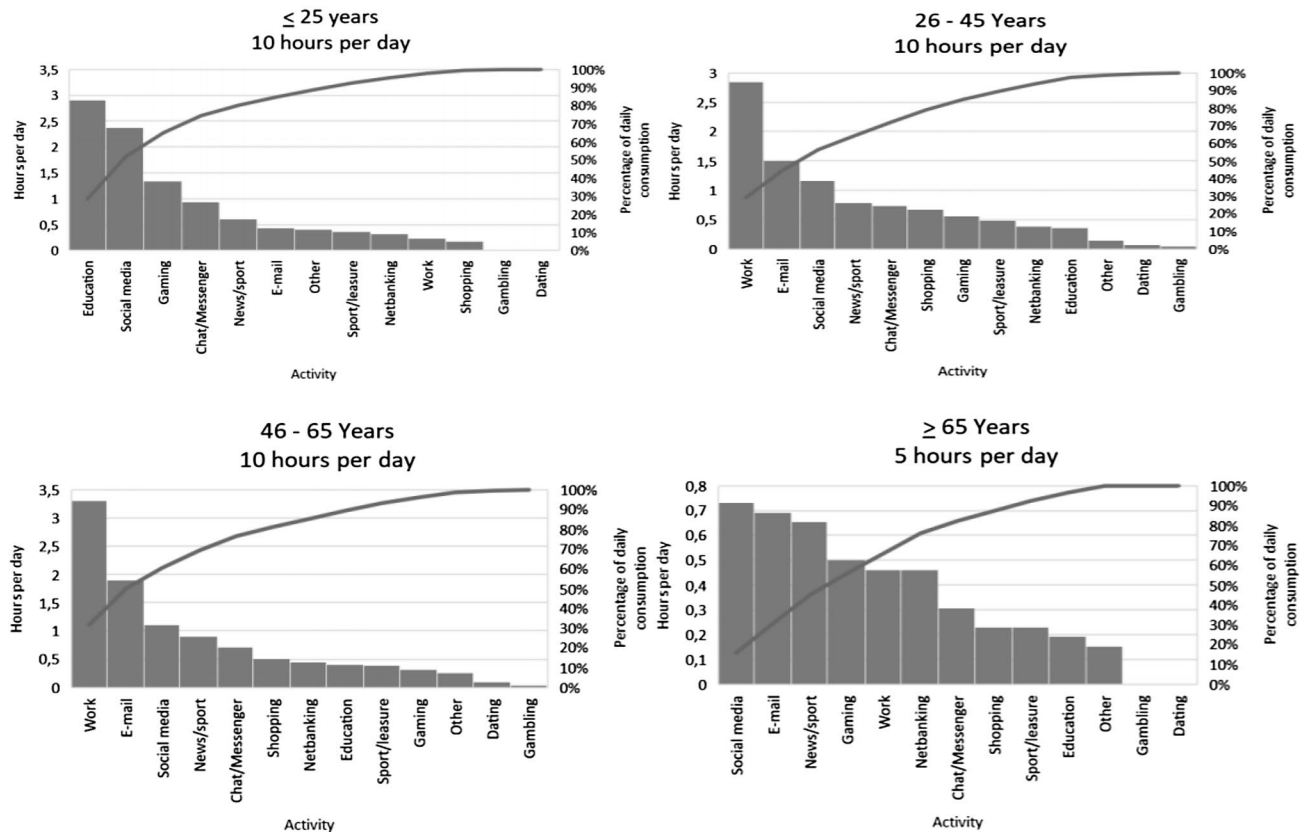

Exhibit 2 Sum of Internet activity per class (age group). *Source* Own elaboration

Table 4 Tabulation of Data-owners' profile with self-protection endeavors

Factor	Gender	Age	Relative frequency (*f*)				
			Never	Biannual	Monthly	Weekly	Daily
A. Reading privacy settings/cookie terms	0.07	0.01	0.74	0.06	0.1	0.04	0.06
B. Changing cookie settings	0.16	0.23	0.64	0.06	0.1	0.1	0.09
C. Opt out (privacy policy unacceptance)	–	− 0.02	0.64	0.07	0.1	0.13	0.06
D. Deleting browsing history	0.13	− 0.01	0.41	0.11	0.25	0.13	0.1
E. Using incognito mode	0.27	0.27	0.58	0.06	0.07	0.15	0.14
F. Blocking website	0.12	0.08	0.63	0.12	0.14	0.06	0.05
G. Manually scanning for malware	0.11	− 0.03	0.61	0.14	0.18	0.07	0.01
H. Automatically scanning for malware	0.19	− 0.09	0.18	0.05	0.09	0.19	0.49

Source Own elaboration

utilized entail the privacy policy (A) and the associated practices pertaining to cookies (B and C.). 74% of the respondents never or almost never read the cookie policy of a website, and 64% will never or almost never alter cookie settings or opt out of using a website due to its privacy policy. The protective measures most utilized, respectively by 0.82 and 0.59 of the respondents are automatic scanning (malware) and the deletion of browsing history. As to the regularity of action, the browsing history is mostly taken on monthly basis. However, the most utilized protective measure, is not a "real" endeavor of self-protection since its application is passive (option H), as the scanning for viruses uses an automated feature of an anti-virus program. Looking at the correlation between the protective measures and gender it appears that in general there is a very weak positive correlation (0.11 to 0.27) on the majority of the protective measures indicating that women apply the measures to a lesser degree than men do. With respect to age, it appears to be a weak positive correlation (0.23 and 0.27) indicating that self-protection endeavors, as to the modification of cookie settings and the use of incognito mode increase with age. Furthermore, respondents exhibited an average of 2½ internet- enabled device/respondent, similar to Gartner's (2019) assessment on IoT devices/individual by 2020 (Gartner 2019).

In sum, data-owners' vested interest, at present, in changing the *status quo,* to extend the privacy over personal data. They find it inconvenient to have others to access their data, and furthermore, they seemed disturbed by the fact that someone is yielding from internet navigation records; most specifically exploiting them for producing marketing campaigns and/or other online ads with a commercial content. However, such marked negative opinion (awareness and sentiment) of individuals is contrary to the current level of action taken to inhibit others from access their data.

From the eight mechanisms of self-protection enquired about, only one of them (automatic scanning for malware) revealed a clear utilization by the majority of the respondents (3 in 4). Noteworthy is, this is a passive self-defensiveness mechanism. In this sense, the "real" endeavors (action) to self-protection of digital footprints seems short, merely circumscribed to one single option (deleting browsing history) and taken merely by half of the respondents on a regular basis.

However, the future of personal data protection it does not seem to change, as the Danish consumers here enquired seem skeptical (less secured) about how their data will be processed. They denote a lower faith on the compliance supervision to compensate their own inaction. Against the current, the gender correlation of women respondents (in IQ#11 and 15), which hold the most negative perception over target advertising, are also the (only) sub-set of the sample willing to incur costs for securing their privacy. This provides a new insight for the private sector, namely for companies on business intelligence, data science and cybersecurity services. Indeed, for marketing professionals within these sectors as well. Moreover, this opens furthermore the horizon of discussion for further research about the privatization of data anonymity services.

(Data-brokers' qualitative) manifest content and Thematic Analysis (TA)

As to the comprehension of the data-broker 'side (perception and action), a thematic analysis (TA) method is applied, as to the exercise of fashioning discrete theoretical themes (codification design) and the subsequent use of a *Gestalt* analysis for unraveling the meaning of the content verbalized by the informants (i.e. the *signifiers*). The latter corresponding to a coding and interpretation of the evidences (quotation) of the coding exercise. Thus, TA implies the establishing of associations of those meanings (i.e. *signifieds*—deriving from *signifiers*) into theoretical foundations fitting into prior acknowledged phenomena, which assumes a recognition of the researcher/s of patterns in transcripts' data, as the coding evidences of these conceptions (Braun and Clarke 2016; Buetow 2010). Hence, our instrumentalization of

TA analysis is rather centered on *primary saliency,* as the linkage of launched signifiers and into realized signifieds (Buetow 2010).

An open coding procedure aiming at illustrating each theme, made use of a situational analysis tool, i.e. the well-known *SWOT* framework (Madsen 2016). Each component (*S–W–O–T*) is equaled to a theme. *Theme 1 (T1)* corresponding to the organizational strengths of the firms F1 and F2 (actions) utilizing *BD analytics resources* (BDAR). *T2* emphasizing the organizational weaknesses regarding its exploitation (i.e. BDAR), including verbalized inactions, indecision or omissions. *T3*, the product-market opportunities for further applicability of BD to consumers as possible innovation routes (awareness and sentiment: data-brokers). *T4* as the market-related weaknesses to its implementation, whether commercial, ethical or legal debilities (perception of awareness and sentiment: data-owners).

In this context, it is relevant to clarify that the explicit content, as sum of the signifiers per participant, accounts a prompting divided into three topics, discussed in this following order: (i) Firm's use of Big Data analytics; (ii) perception of BD's benefits/drawbacks for the firm; and, (iii) perception of benefits/drawbacks to the consumer. Such design reflects an introductory discussion of BD morphology and future traits, followed by an immersion on the firm's seizing (thoughts versus actions) of BD benefits; and, finally, the understanding of the consumer sentiment, as to their willingness to be exposed and accordant endeavors (Table 5).

The open coding of the (qual) data-brokers' interviews unraveled 10 signifieds with the majority (0.6) covering threats, although, with the second largest (0.3) theme covering opportunities. As to the strengths (T1) is pinpointed the development of new product solutions, refinement of the portfolio, service support and feedback systems towards the consumer. T3 highlighted the (un)know experience gains, as to the correction of defects, improvements of technology with direct expression in ambient and effort/energy-saving for the consumer. Furthermore, other opportunities acknowledge the benefits of BD's economies of scope through the subsequent exploring of collaborate advantages regarding cross-enterprise cooperation. However, the latter shields the other side of the coin, as to the potential exploitation of these economies of scope across firms/industries against the consumer well-being. Thus, regarding the T4, the participant F1, referred to the real risks of misuse of the BD ecosystem for enhancing collaborative advantages. In fact, has provided a virtual example of such hazard: "…imagine an American fast-food retail giant partnering with a Danish biotech company whom produces blood pressure measuring devices and together monitoring sugar level fluctuations, suggesting their consumers, in real-time, to buy their burgers/pizzas, this is a harmful use of data…" (Quotation 2—T4:2:1:2; of CId T4:2:1).

Furthermore, the divergent regulations across the globe were asserted as hampering a wider and faster spreading of strengths and opportunities associated with BDAR. Another identified threat was the current degree of confidence (or belief) of the consumer on search engine optimization (SEO) ranking and reputation management software, rating systems and the perceived consumer peer reviewing experiences. Such services delivered among others by providers such as, *Apple Store; BirdEye; Podium; TrustPilot; Yotpo;* or *Google Play*, were portrayed as holding the power to influence data-sharing willingness, contributing therefore to polarization between acceptable/unacceptable standards, high/low ratings, with an inherent loss to data-brokers data inflow and

Table 5 Coding data outputs per theme

Theme	Code		Areas of intervention
	CId*	Description	
T1: S-action	*T1:1:1*	Product/s improvement	New product develop. (NPD)
T2: W-action	–	–	–
T3: O-sentiment**	*T3:1:1*	Service optimization—experience	Customer experience (CX) transformation
	T3:2:1	Service optimiz.—health/well-being	Health improvement (HE)
	T3:3:1	Collaborative efficiency (benefit)	Explore cooperation strategies (CS)
T4: T-sentiment**	*T4:1:1*	Informal data breach	Data-shared to personal networks
	T4:2:1	Collaborative efficiency (hazard)	Exploit cooperation strategies (CS)
	T4:3:1	BD's divergent legal framework	Hamper the S–O's dissemination
	T4:4:1	Overvalue of rating systems	Customer experience (CX)
	T4:5:2	Data pollution	BD software data fluxes
	T4:6:2	GDPR compliance	BD's legal framework

Source Own elaboration

*Cid—Code's identification

**Sentiment equaled to "perception" on the research framework

data-owners delivered value. The biased data inputs returned by these type of software applications is claimed to be a matter of managing *data pollution*, as labeled as garbage-in determining garbage out.

Participant 2, considered to be furthermore "….as in an arms race…" especially against North American and Asian companies, with the GDPR being costly and a time-consuming limitation. The GDPR hinders the development of *Artificial Intelligence* (AI) of EU firms, obliged to restructure and fully comply in terms of their data collection, handling, and the storage infrastructure (including their destruction policies). Consequently, both participants perceive legislation as the biggest threat to the exploitation of BD. Yet, both companies assert Big Data as "need" for their survival, and a "stick" behind the necessity.

Triangulation (analytical) procedure

As indicated in section methodology, this study adopts a data triangulation and data-theory triangulation procedure to achieve a better picture of the factual and the real (Altricher et al. 1996). This occurs for a dyadic purpose. First, to validate primary data obtained from data-brokers and data-owners in the light of a "sense-giving" analytical procedure (data triangulation). Secondly, to extend our findings and deliver a higher contribution both to practitioners in the industry (senior/middle managers) and to academics whom may be also conducting research within the Big Data realm (data-theory triangulation). This instrumentalization of the triangulation procedure is compliant with the universal practices for cross-examination of heterogeneous sources, whether neoclassic or contemporary ones, pursued by multiple methodologists (Turner and Turner 2009; Lincoln and Guba 2000; Altricher et al. 1996).

Regarding data triangulation, we have revised the TA method qual outputs of the data-brokers with the quants applied to the firm's consumers. Recalling the type 3 variables (behavior) of defensiveness, *var003-2* (data-owner's self-protection) referring to RQ3 and O2.1, reveals a low degree of endeavor towards personal data protection, with the respondents figures (0.74 and 0.64) indicating that a majority of them have never used the factor A ("Reading privacy settings/cookie terms") and factor C ("Opt out (privacy policy unacceptance)"). For the validation of the prior we have used qual data from the data-broker's side. Here, Participant 1 signifier is inconclusive but the Participant 2 corroborates the results above acknowledging a literal null opposition of their customers to the firm cookies policy:"

> All cookie responses are monitored, and the current status is, that 100% of the visitors to the websites give consent to cookies and thereby to sharing personal data. There have thus far not been any complaints from customers of [Firm 2] in relation to its privacy policy.

Yet, we ought to emphasize the incongruence (*action vs. perception*), since the perception denotes a 70% of the respondents declaring in IQ# *11* to have a negative sentiment about their personal data utilization for commercial purposes. Furthermore, this is accentuated by the rather coherent answers to IQ#14 which confirm such negative sentiment as more than half (0.58) of the sampled individuals are willing to pay a short fee for maintain their privacy.

As to respondent's age significance, the tabulation of self-protection by age per factor (A; C) exhibited respectively a very weak positive, and a very strong negative relation, respectively. However, results are inconsistent when crossed with the signifiers of the qual method, since *Participant 1* expressed a dissimilar perspective, arguing on the existence of two main consumer profiles, which he calls category A—*digital natives*; and category B—*digital immigrants*, with a perceived different sentiment toward their digital footprints. The *digital natives* (category A), refer to the generation z and millennials, those who fit in the sample onto one age group (under 25 years old) meaning that they are born or brought up during the age of digital technology and this participant considers they accept more easily to share data with internet service providers. Whereas, the category B (digital immigrants) experience most reluctance in accepting to share data. Those are the ones the participant termed them as "unwise" since they are born or brought up before the widespread use of digital technologies, they are more resistant in sharing data, thus more aware of such implications, but nevertheless, accept generally the terms and conditions without reading or understanding them.

Furthermore, we have conducted a cross-observation of quants/qual data outputs from the aggregate of primary data alongside with the theoretical review conducted in "Literature review" section (data-theory triangulation). This procedure is though a hybrid one, accounting both a pure data triangulation and pure theoretical triangulation, in which the latter represents the use of multiple concepts, as subthemes within the field, gathered from the theoretical revision to confirm its observation in the empirical testing phase (Turner and Turner 2009; Dzurec and Abraham 1993). This data-theoretical procedure is as postulated by other methodologists an "attempt to map out, or explain more fully, the richness and complexity of human behavior by studying it from more than one standpoint" (Cohen and Manion 1986, p. 254). Table 6 below deliver a triangulative output for further interpretation and so building a broader meaning system as to the perception *versus* action of both parts.

The results above highlight the mapping of 42 implications of digital footprints (for data-brokers and data-owners) as to the utilization of Big Data Analytics' resources

Table 6 Qual Data-theory (BDAR) triangulation

Theme	Sub-theme (ST_n)	Data (Qual) Sub-theme	Implication	Sub-theme (ST_n)	Theory ("Literature review" section) Sub-theme	Implication
T1: S	T11S	Product improvement and NPD	Data-brokers Data-owners	T12S	Access endless sources/entertainment	Data-owners
				T13S	Convenient access (to data)	Data-owners
					Experience empowerment	Data-owners
				T14S	Low resource consumption (e.g. time)	Data-owners
				T15S	Free-recommendation/counseling	Data-owners
				T16S	Optimization of satisfaction	Data-owners
				T17S	Short-cut to desired goods	Data-owners
T2: W		–	–	T21W	Data fumes and RTBF	Data-owners
				T22W	Digital voyeurism	Data-owners
				T23W	Price discrimination	Data-owners
				T24W	Price steering	Data-owners
				T25W	Streisand effect	Data-owners
				T26W	Hyper-targeting (micro-targeting)	Data-owners
T3: O	T31O	Collaborative efficiency	Data-brokers	T350	Tech-developments (Digisense, IoT, AI)	Data-brokers
	T32O	Customer experience transformation	Data-brokers Data-owners	T36O	Incident prevention, safety, security & health lifting	Data- brokers
	T33O	Service optimization—experience	Data-brokers Data-owners	T37O	Governability & Law enforcement	Data- brokers
	T34O	Service optimization—health/well-being	Data-brokers Data-owners	T38O	Economic and business development (2nd economy; 4th industrial revolution (IR)	Data- brokers
				T39O	New Product Development (NPD)	Data-brokers Data-owners
				T310O	Development of authorities' public intelligence	Data- brokers
				T3111O	Scientific knowledge advancement	Data- brokers Data-owners
T4: T	T41T	Collaborative-advantage's hazards	Data-owners	T47T	Consumer's controlling (meta-preferences)	Data-owners
	T42T	Data breach	Data-brokers Data-owners	T48T	Consumer welfare depreciation	Data-owners
	T43T	Data pollution	Data-brokers Data-owners	T49T	Data breach	Data-owners
	T44T	Divergent legal frameworks	Data-brokers Data-owners	T410T	Dataveillance	Data-owners
	T45T	GDPR's compliance (EU firms)	Data-brokers Data-owners	T411T	Gap regulation vs. inspection (GDPR)	Data-brokers Data-owners
	T46T	Overvalue of rating systems	Data-brokers	T412T	Level of individual capabilities to self-protection	Data-owners
				T412T	Mind-reading & manipulation	Data-owners
					Personality profiling	Data-owners
				T413T	RTBF ineffectiveness (GDPR)	Data-owners
				T414T	User reactiveness	Data- brokers
				T415T	Sentiment manipulation	Data-owners

Source Own elaboration

for commercial purposes. From the absolute frequency, 3 of them co-occur ($f = 0.07$) between the two components of data/theory (i.e. data breach; GDPR; NPD). The theoretical-driven mapping seems dominant over then data-driven mapping, with the first corresponding to 76.19% of all encountered implications.

Noteworthy is also that the impact being mostly felt on the data-owners' side, since 64.15% of the typology from the total of 42 implications observed, whether positive or negative falls on data-owners' side. However, 12 units (22.64%) yield an impact on both data-owners and brokers.

A positive relation *(S1: strengths)* is conveyed by 8 sub-themes with full implication on individuals, and a more incipient figure (0.125) on the data firms, solely affecting the latter as to the ability to improve products and NPD). The current negative one *(T2: weaknesses)* contain also 8 sub-themes all affecting solely the data-owners. These units came solely from the theoretical revision since the coding of the interviews to the participants (as data-broker's

Table 7 Perception of digital footprint's implication per societal domain

Domain	Sub-theme		T	Domain's freq. (*f*)			
	Id	Description	Id	S	W	O	T
Economic (Econ)	T38O	New Business development	T3–O	6	3	6	7
	T47T	Consumers' controlling (meta-preferences)	T4-T				
	T412T	Consumer's manipulation	T4-T				
	T415T		T4-T				
	T414T	Consumer defensive reactionism	T4-T				
	T32O	Consumer satisfaction	T3-O				
	T12S		T1-S				
	T12S		T1-S				
	T15S		T1-S				
	T16S		T1-S				
	T17S		T1-S				
	T48T	Consumer welfare depreciation	T4-T				
	T26W	Hyper-targeting (micro-targeting)	T2-W				
	T31O	Networking and Collaborative advantages	T3-O				
	T39O	New Product Development (NPD)	T3-O				
	T46T	Rating systems' bias	T4-T				
	T23W	Price discrimination	T2-W				
	T24W	Price Steering	T2-W				
	T11S	Product innovation	T1-S				
	T33O	Service optimization	T3-O				
	T34O	Service optimization	T3-O				
	T25W	Streisand effect	T2-W				
Legal (Leg)	T410T	Dataveillance	T4-T	–	1	–	7
	T44T	Heterogeneity of legal acts worldwide	T4-T				
	T412T	Self-defensiveness (GDPR)	T4-T				
	T45T	GDPR Compliance	T4-T				
	T411T	Gap regulation vs. inspection (GDPR)	T4-T				
	T21W	GDPR individual rights (RTBF)	T2-W				
	T37O	Inspection/Law enforcement	T3-O				
	T413T	RTBF ineffectiveness (GDPR)	T4-T				
Medical (Med)	T36O	Incident prevention, safety, security & health lifting	T3-O	–	–	1	–
Technological (Tec)	T42T	Data breach	T4-T	–	2	3	3
	T49T		T4-T				
	T21W	Data Fumes	T2-W				
	T43T	Data pollution	T4-T				
	T22W	Digital Voyeurism	T2-W				
	T35O	Tech-developments	T3-O				
	T310O	Development of public intelligence	T3-O				
	T311O	Scientific knowledge advancement	T3-O				

Source Own elaboration

representant) did not generate any quotations, as perceived evidence of this phenomenon.

As to the perception of the future opportunities brought by these digital footprints (*T3: opportunities*) 11 sub-themes emerged with implications mostly felt over the data-brokers side, as to ¾ of benefits. The risk of future losses *(T4: Threats)* uncovered 18 sub-themes or potential implications, here affecting dominantly the data-owner's side (0.8333) (Table 7).

The large majority of implications (or sub-themes) fall within the economic landscape, which covers over one-half of the evidence collected from the empirical exercise (including the triangulation with the theoretical revision). Although, the legal and technological landscape are relevant domains of influence for Big Data analytics resources and digital footprints' exploration, accounting for 0.225 and 0.2, respectively. Furthermore, considering these outputs, we have modeled the *impaction index* (II) of BD analytics and digital footprints, considering the relative impact (RI) of each theme (T1 to T4).

$$\mathrm{II}_{i=\mathrm{Sum}}(\text{current positive implications} / \textit{overallpositiveimplication}) + \Delta(T_n) \quad (6)$$

$$\mathrm{II}_{i=}\left(\mathrm{RI}_{(\text{data-brokers})} U\ \mathrm{RI}_{\text{data-owners}}\right) + \Delta_{t=1}\left(\mathrm{RI}_{\text{data-brokers}} U\ \mathrm{RI}_{\text{data-owners}}\right) \quad (7)$$

Considering furthermore that the typologies of impact (S–W–O–T) are by definition identical, where as

$$\mathrm{RI}_{i\,(\text{data-brokers})} \equiv \mathrm{RI}_{i\,(\text{data-owners})}$$

$$\mathrm{RI}_{i\,(\text{data-owners})} = \sum\left(\left(\frac{S}{\mathrm{OS}} - \frac{W}{\mathrm{WT}}\right) + \left(\frac{S}{\mathrm{OS}} - \frac{W}{\mathrm{WT}}\right)\right) \quad (8)$$

While the variation function, considers the arithmetic combination of the impact function in present ($t=0$) and in future ($t=1$) here represented in the polynomial expression.

$$f(x) = \mathrm{SWOT}_{(t1)} - \mathrm{SWOT}_{(t0)} \quad (9)$$

The computing of these results of strengths (theme 1) over weaknesses revealed a shared benefit exploitation for both data-owners (1.33) and data-brokers (1.00). Conversely, the perception as to the present/future revealed dissimilar results. For the first, the opportunity over threats function is clearly a negative one (0.33) contrasting with the latter ones (1.57). The impaction index ratios *(0,9011* and *1,57,143)* denoted a better exploitation function of BD for the firm's side than the individuals.

Discussion

We have deduced from the empirical testing presented in "Data analysis and discussion" section that the individual benefits of the digital exposure of personal data came at the expense of the legal and economic consequences of the consumer. On the legal landscape, this encompassed the unprotecting of legal rights (namely the ones consecrated on the GDPR described in "Literature review" section) and the abusive utilization of data by third parties. Thus, consumers experience the non-safeguarding of their rights and misuse of unauthorized data. On the economic landscape, consumers are being targeted, manipulated and deceived in current and future product-offerings, offer delimitation and price discrimination for the optimization of corporate profits.

Data-brokers have also benefited from BD analytics resources (BDAR), modifying and/or extending portfolios and refining digital positions. The major drawbacks of BD analytics, pinpointed at the interviews, is the creation of data pollution. We argue that the economic consequences of digital exposure came as the tail of the legal landscape, especially due to the debilities discussed in previous sections of the compliance supervisory system.

However, the participants signaled relevant information as to the future threats of BD, both for corporations and consumers. For instance, the overvaluing of rating systems and its inherent bias and/or deception as fake data. These constitute an example of what might be a faulty use of data fumes, which can in turn, manipulate the consumer and polarize the preferences of products/services in favor of large and dominant incumbents. This is a phenomenon which one participant pinpointed as a type of data pollution. The participant claimed also data pollution has spread to multiple quadrants of our society and is used with multiple purposes. He argued that the spiraling up of this phenomenon is yet unpredictable. Here, another participant stressed how unprepared organizations are nowadays to deal with informal data breach, which may even make more complex even more than current typology, flow and directionality of BD in circulation. Yet, also opportunities to data-brokers were clearly manifested, as to cooperation, networking and from these extract collaborative advantages. Also, the room for service optimization brought by tech-developments on the new digisense era of artificial intelligence, machine learning and IoT and their application to the industry (4th industrial revolution—or industry 4.0) with inherent benefits for broad scientific developments as well.

The perceptions and actions revealed though however a clear gap emphasized by the most mature age groups. Consumers denote some *laissez-faire* and a passive acceptance of non-agreed practices of data-brokers. Results from self-protection, as the enquired summarized in Table 4. *Tabulation of Data-owners' profile with self-protection endeavors*, demonstrated a very low extent of proactive defensiveness (from the range of given options—from A to H). These options were incipiently used, and the most frequent being an automated one (a computer automatic check of malware). Thus, the generalized negative perception of Big Data among consumers appears disproportionate to the protective measures adopted by the same consumers to mitigate possible undesired effects of dataveillance and analytics practices. Data-owners denoted in general some apathy and resignation adopting mostly passive-protection measures, as the most significant being automatic scanning of malware.

Furthermore, the respondent's open- comments conveyed in comments boxes revealed that consumers seem to rely faithfully on GDPR protection rather than self-protection, a phenomenon also acknowledged before by Baruh and Popescu (2017). Women respondents exhibited though the desire to alter the current context, revealing a higher predisposition for spending on small payments on internet services to become problem-free of commercial exploitation of personal data.

Finally, it shall be emphasized though that both collection methods and analytical procedures that allowed us to map a broad range of implications of BD, were discussed in number (or breadth) but not in the degree (or depth) of influence inflicted on the consumer or the company. Thus, we recommend other researchers in the field to explore the results of this project on the Danish market and replicate them to other Nordic countries and furthermore pursue new avenues for enlightening the extent of influence per theme and compare them through replication on multiple research angles.

Conclusions

The initial euphoria surrounding Big Data has clearly vanished. Two decades ago, the marvel of the digital markets and online platforms have attracted many firms to enhance their competitive positioning through the electronic channels plus to re-segment towards other consumer's targets, subsequently reaping the economic benefits of such strategic decisions. Online businesses assisted a steady growth against brick-and-mortar shopping, while the commoditization of the web through the user's daily usage has, with time, depreciated the technological benefits. Firms in most digital societies (as Denmark, here in examination), feel an accentuated market pressure for the attraction of attention of the consumer and for the efficient usage of their data. As referred by one of the participants in this study, it is a "… arm race…". Big Data is a "need" and companies identified an element of "stick" behind this necessity.

The results of our investigation on two large technological firms on the Danish market are consistent with the revised literature. Both firms and consumers seem discontented as to the current state of the EU's data protection policy. Consumers have turned their attention to the safeguarding of individual rights over personal (profiling) data and navigation/ consumption footprints' data. The major discontent of data-owners can be traced to a couple of issues. Firstly the "Right To Erasure" or Right to Be Forgotten" (RTBF). Secondly, the (alleged) predatory practices for exploiting third parties' data. Consumers assert to be hyper-targeted and price discriminated/steered. Firms claim that global competition is tight and policies are a competitive limitation for European firms, and especially restrictive for advancements in artificial intelligence.

The perception of personal data utilization (*objective 1—O.1*), is not accompanied by according actions (O1.1.) prevailing the hazard of *laisser faire*. For instance, the category "sentiment" within the perception category (*var002-2—exposure willingness*) does not come together with the defensiveness behavior (*var003-2 self-protection*) of data-owners to secure their own privacy. For instance, the tabulation of these categories (sentiment with defensiveness) determined that the sole significant mechanisms of self-protection is, in vast majority, a passive mechanism (i.e. automatic malware scanning). Noticeably, the unique action of self-protection of approximately ½ of the respondents is the regular monitoring of the browsing history. As to the data-brokers' side *(objectives 2 and 2.1.-O2; O2.1.),* data revealed an equivalent gap of perception versus action. Yet, we argue that companies' myopia (perception) or isomorphism (inaction) seems to be clouded by the competitive needs of their businesses.

The bottom-line issue here uncovered by such perception–action gaps is that the consumer's expected degree of privacy is significantly higher than the action taken to ensure the first. Moreover, data-owners seem unable to secure their own rights (RTBF) per se, and moreover, are not endowed with technical monitoring tools or mandated with executive power to do so. Finally, the role of self-supervision seems a burden for the individual (even demonstrating a clear negative perception of privacy, relying faithfully on a better future as to the GDPR's Compliance Supervisory Model.

Scholars in earlier literature have identified a mixture of major causes for the negative perception over Big Data analytics which can be summarized as follows: (i) insufficient regulation; and, (ii) debilities on the current legal framework (i.e. the GDPR). These causes are argued to be the root for the proliferation of deviating and unethical practices. Scholars point in special the finger to the articles 17 and 37 of the

GDPR, asserted to have failed to safeguard the individual data protection rights and the competitiveness of European firms; but also, targeting the supervisory system conferred to the European Data Protection Supervisor (EDPS) (Baruh and Popescu 2017; Xue et al. 2016; De Mauro et al. 2015; Hannak et al. 2014; Thatcher 2014).

In this context, we do concur with Baruh and Popescu's (2017) assertion that this problem is holistic, and is a matter of failure of the regulatory system. The current legal framework encountered in the GDPR from 2018 is entirely built on the assumption that "privacy" is an individual intangible good, which is therefore justified to be secured by the skills of the person/owner (individual capabilities) to the self-administration of the asset. These authors reminded us though that privacy is, not an individual, but a phenomenon of collective value for the whole social regime, which is supported by the article 5's principles of lawfulness and fairness, and is furthermore, a fundamental right defined on the article 8(1) of the Charter of Fundamental Rights of the European Union. Hence, the definition of simply directional policies seem sparse, such as, in article 37, the imposition to appoint an internal Data Protection Officer (DPO), or nominate an external one, as policies lack of specifics as to spectrum of responsibilities, and elucidation of the seriousness of a faulty practice. It is our understanding that there is still categorically a road to be traveled in Denmark and in the EU regarding the regulatory affairs on personal data (including digital footprints).

We advocate a reform of the current GDPR's Compliance Supervisory Model, to reduce the burden of self-protection and enlarge the operational span of control of the regulator, currently limited in resources to an advisory organ (the European Data Protection Board) and a supervisory organ (the EDPS).

In parallel, firms hold undoubtedly the responsibility of complying scrupulously with legal conventions, reversing reputational and economic risks of past practices while exploring data on a risk-free manner. This means, to avoid data breaches and administrate transparently the easy-to-use access to data. Data-brokers, are recommended to build IT-related dynamic capabilities (namely on data science and cybersecurity) in order to explore, on a positive manner the window of opportunity given by BD analytics. The rationale is less imitation and more innovation. Here is important to focus on the gauging of the consumer's sentiment and attend to the willingness of particular sub-segments to experiment new ideas and concepts, such as, the sub-set of respondents in our sample, whom signaling their openness to paid for data protection services (presented on "Data-owners" section).

It is though imperative that the refashioning of the regulatory system can carry regulation and stronger inspection mechanisms on collective protection, cover current legal gaps on BD analytics' practices (e.g. data fumes or *Streisand effects)* and make a more equitable redistribution of accountabilities between authorities, firms and individuals. Negatives practices of algorithmic refinement for the predictive modeling ought to be more tightly scrutinized and not left to the self-arbitration of industry practitioners. Finally, the meta-national supervision hub (EDPS) is suggested to expand the interaction with EU State-Members' authorities and share a larger extranet of both joint technical expertise, but also, share the responsibilities of inspection.

Declarations

Conflict of interest The authors have no conflicts of interest to declare.

References

Abrantes, B.F., and A. Venkataraman. in press. Environment kinesis and organizational adaptability: Effects of EU's general data protection regulation (GDPR) on the Danish software industry. *International Journal of Learning and Change*. https://doi.org/10.1504/IJLC.2020.10033872.

Abrantes, B.F. 2020. Tech-innovation and spillovers on corporate-defensiveness: Evidence from the Lisbon startup ecosystem. *International Journal of Business Competition and Growth* 7 (1): 68–100.

Ahn, J., and J. Lee. 2020. Case study on big data sampling population collection method errors in service business. *Journal of Service Research and Studies* 10 (2): 1–15.

Altrichter, H., P. Posch, and B. Somekh. 1996. *Teachers investigate their work: An introduction to the methods of action research*. London: Routledge.

André, Q., Z. Carmon, K. Wertenbroch, A. Crum, D. Frank, W. Goldstein, J. Huber, L. Van Boven, B. Weber, and H. Yang. 2018. Consumer choice and autonomy in the age of artificial intelligence and big data. *Customer Needs and Solutions* 5 (1–2): 28–37.

Barlow, J.P. 1996. Declaration of independence for Cyberspace, February 8th 1996. Retrieved from https://www.eff.org/cyberspace-independence. Accessed 5 April 2020.

Baruh, L., and M. Popescu. 2017. Big data analytics and the limits of privacy self-management. *New Media & Society* 19 (4): 579–596.

Benjelloun, F.Z., A.A. Lahcen, and S. Belfkih. 2015. An overview of big data opportunities, applications and tools. In *2015 Intelligent Systems and Computer Vision (ISCV)*, pp. 1–6

Boyd, D., and K. Crawford. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15 (5): 662–679.

Braun, V., and V. Clarke. 2016. (Mis)conceptualising themes, thematic analysis, and other problems with Fugard and Potts'(2015) sample-size tool for thematic analysis. *International Journal of Social Research Methodology* 19 (6): 739–743.

Brayne, S. 2017. Big data surveillance: The case of policing. *American Sociological Review* 82 (5): 977–1008.

Buetow, S. 2010. Thematic analysis and its reconceptualization as 'saliency analysis.' *Journal of Health Services Research & Policy* 15 (2): 123–125.

Cavanillas, M.J., E. Curry, and W. Wahlster. 2016. *New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe*. New York: Springer.

Chen, H., R.H. Chiang, and V.C. Storey. 2012. Business intelligence and analytics: From big data to big impact. *MIS Quarterly* 36 (4): 1178.

Cohen, L., and Manion, L. 1986. *Research Methods In Education*. London: Croom Helm.

Cox, M., and D. Ellsworth. 1997. Managing big data for scientific visualization. *ACM Siggraph* 97 (1): 21–38.

Crawford, K., and J. Schultz. 2014. Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review* 55 (1): 93–128.

De Hert, P., V. Papakonstantinou, G. Malgeiri, L. Beslay, and I. Sanchez. 2017. The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review*. https://ac.els-cdn.com/S0267364917303333/1-s2.0-S0267364917303333-main.pdf?_tid=c61e5bb2-11b8-464d-98a2-f01508a953cf&acdnat=1520597308_0f87eb7bf694fb7eeb6d95f7f20054df.

De Mauro, A., M. Greco, and M. Grimaldi. 2015. What is big data? A consensual definition and a review of key research topics. *AIP Conference Proceedings* 1644 (1): 97–104.

Dekimpe, M.G. 2020. Retailing and retailing research in the age of big data analytics. *International Journal of Research in Marketing* 37 (1): 3–14.

Demant. 2019. 2019 Annual report, nd, https://www.demant.com/investor-relations/annual-report-2019. Accessed 7 May 2020.

Diamond, L. 2019. The road to digital unfreedom: The threat of postmodern totalitarianism. *Journal of Democracy* 30 (1): 20–24.

Dodge, M., and R. Kitchin. 2005. Code and the transduction of space. *Annals of the Association of American Geographers* 95 (1): 162–180.

Dzurec, L.C., and I.L. Abraham. 1993. The nature of inquiry: Linking quantitative and qualitative research. *Advances in Nursing Science* 16 (1): 73–79.

EU. 2020. EU data protection rules, https://ec.europa.eu/info/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules/eu-data-protection-rules_en#documents. Accessed 16 March 2020.

Eur-Lex. 2016. Regulation (EU) 2016/679 of the European Parliament and the Council of 27 April 2016, Official Journal of the European Union, L119/1. https://eurlex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02016R0679-20160504&from=EN. Accessed 5 April 2018.

European Data Protection Supervisor (EDPS). Online. The History of the General Data Protection Regulation. https://edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en. Accessed 25 March 2021.

Evans, J.D. 1996. *Straightforward statistics for the behavioral sciences*. Washington: Thomson Brooks/Cole Publishing Co.

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

Gartner. 2019. Gartner Glossary, nd, https://www.gartner.com/en/information-technology/glossary/big-data. Accessed 5 Dec 2019.

Hannak, A., G. Soeller, D. Lazer, A. Mislove, and C. Wilson. 2014. Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 conference on internet measurement conference*, 305–318.

Hilbert, M., and P. López. 2011. The world's technological capacity to store, communicate, and compute information. *Science* 332 (6025): 60–65.

González, R.J. 2017. Hacking the citizenry? Personality profiling, 'big data'and the election of Donald Trump. *Anthropology Today* 33 (3): 9–12.

IDC. 2018. *5 Things You Didn't Know About Tech Spending*. https://blogs.idc.com/2018/10/11/5-things-you-didnt-know-about-tech-spending/. Accessed 8 Dec 2019.

Information Commissioner's Office (ICO). 2018. *Guide to the General Data Protection Regulation (GDPR)*. Paper 1.0.248. https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf. Accessed 25 March 2021.

Jacobs, A. 2009. The pathologies of big data. *Communications of the ACM* 52 (8): 36–44. https://doi.org/10.1145/1536616.1536632.

Jones, T. M. 1991. Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of management review* 16 (2): 366–395.

Kim, G.H., S. Trimi, and J.H. Chung. 2014. Big-data applications in the government sector. *Communications of the ACM* 57 (3): 8.

Kitchin, R., and G. McArdle. 2016. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*. 3 (1): 1–10.

Lambrecht, A., and Tucker, C.E. 2015. Can Big Data Protect a Firm from Competition?. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2705530.

Leinweber, D.J. 2007. Stupid data miner tricks: Overfitting the S&P 500. *The Journal of Investing* 16 (1): 15–22.

Lecuona, I., and M. Villalobos-Quesada. 2018. European perspectives on big data applied to health: The case of biobanks and human databases. *Developing World Bioethics* 18 (3): 291–298.

Lerman, J. 2013. Big data and its exclusions. *Stanford Law Review Online* 66: 55.

Lincoln, Y.S., and E.G. Guba. 2000. Paradigmatic controversies, contradictions, and emerging confluences. In *Handbook of qualitative research*, ed. N.K. Denzin and Y.S. Lincoln. Thousand Oaks: Sage.

Madsen, D.Ø. 2016. SWOT analysis: A management fashion perspective. *International Journal of Business Research* 16 (1): 39–56.

Manovich, L. 2011. Trending: The promises and the challenges of big social data. *Debates in the digital humanities* 2 (1): 460–475.

Marz, N., and J. Warren. 2012. *Big Data: Principles and best practices of scalable real-time data systems*. MEAP. Westhampton: Manning.

Marr, B. 2017. *Really big data at Walmart: Real-time insights from their 40+ petabyte data cloud*. https://www.forbes.com/search/?q=Really%20big%20data%20at%20walmart#fbc1599279f4. Accessed 17 Dec 2019.

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

Muhammad, S.S., B.L. Dey, and V. Weerakkody. 2018. Analysis of factors that influence customers' willingness to leave big data digital footprints on social media: A systematic review of literature. *Information Systems Frontiers* 20 (3): 559–576.

Russom, P. 2011. Big data analytics, TDWI Best Practices Report. *Twi Research* 19 (4): 1–34.

Semerádová, T., and P. Weinlich. 2019. Computer estimation of customer similarity with Facebook lookalikes: Advantages and disadvantages of hyper-targeting. *IEEE Access* 7: 153365–153377.

Steinberg, E. 2020. Big Data and Personalized Pricing, Business Ethics Quarterly, vol. 30, No. 1, 97–117, https://econpapers.repec.org/article/cupbuetqu/v_3a30_3ay_3a2020_3ai_3a1_3ap_3a97-117_5f5.htm.

Thatcher, J. 2014. Big data, big questions| Living on fumes: Digital footprints, data fumes, and the limitations of spatial big data. *International Journal of Communication* 8 (1): 1765–1783.

Turner, P., and S. Turner. 2009. Triangulation in practice. *Virtual Reality* 13 (3): 171–181.

Van Dijck, J. 2014. Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society* 12 (2): 197–208.

Xue, M., G. Magno, P. Landulfo Teixeira, E. Cunha, V. Almeida, and K.W. Ross. 2016. The right to be forgotten in the media: A data-driven study. *Proceedings on Privacy Enhancing Technologies* 2016 (4): 389–402.

Zuboff, S. 2019. *Surveillance capitalism and the challenge of collective action*. CA: SAGE Publications.

Zwitter, A. 2014. Big data ethics. *Big Data & Society*, 1–6.

Bruno F. Abrantes is an Associate Professor on the field of Global Strategic Management at Niels Brock Copenhagen Business College (NBCBC) (Denmark) and a member of the Strategic Management Society (SMS). His research interests encompass the fields of dynamic capabilities and international business.

Klaus Grue Ostergaard is a B.A. (Hons) graduate on Business Administration from De Monfort University (UK) and a professional in the field of Defense & Security with research interest for Big Data (Analytics).