Digital footprints´ wrangling – Are analytics used for better or worse?: Evidence of Danish data governance practices on tech-based services

Bruno F. Abrantes (bruno.abrantes.dk@gmail.com)
Copenhagen Business College https://orcid.org/0000-0002-4812-8914
Klaus Ostergaard
Niels Brock: Niels Brock Copenhagen Business College

Research Article

Keywords: Big data, consumer´s sentiment, Denmark, digital footprints, GDPR

Posted Date: June 4th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-472309/v1

Digital footprints´ wrangling – *Are analytics used for better or worse?*: Evidence of Danish data governance practices on tech-based services

Abstract

Incumbents face nowadays one of their great data governance challenges, regarding the balancing of Big Data´s utilization and the imperative of uplifting the consumer´s sentiment. In the context of the European Union (EU), the prior is especially applicable as a growing opposition against the allegedly hazardous commercial exploitation of *digital footprints* is ascribed to distort information bundles to the consumer and gradually bias their individual capacity to take informed free choices, configuring a phenomenon of concealed "data autocracy". Yet, this occurs despite the introduction on May-18 of the *General Data Protection Regulation* (GDPR).

Hence, we have conducted a multi-methods enquire with *foci* on the Danish market covering the footprint´s awareness and undertaking´s rationale targeting equidistantly the data-owners and data-brokers. This was fashioned with a descriptive-exploratory research purpose to enlighten both the sentiment (perceptions) and the behaviour (actions) and determine the extent of (un)willingness surrounding the *dataveillanc*e over personal lives (and firm´s quotidian) and their coherence with regard to self-protection. A dyad of triangulation procedures were applied then for validation purposes and extending our findings.

The results of the empirical testing confirmed the most marked disruption of Big Data at individual-level as to the range of benefits and drawbacks. A gap (perception *versus* action) acknowledged on the two parties uncovered a (paradoxical) *laissez-faire* of data-owners and institutional isomorphism of data-brokers, and the need for capability building and for the government authorities´ intervention, re-conceptualizing scope and boundaries of collection securitization for building a far more efficient model against citizen´s rights usurpation.

Keywords: Big data; consumer´s sentiment; Denmark; digital footprints; GDPR

JEL: M15; M31; M38.

1. Introduction

1.1. Background and initial problematization

The indelible evolution of the social regimes in recent decades leveraged by constant technological advancements, seemed to be, on itself, profoundly shaped in recent decades by the advent of information and communication technologies´ (ICT) disruptions, molding the contemporary world towards a digital era with a significant impact on the organizational sphere and citizen´s lives.

Indeed, the quest for a boost on the organization´s capacity to accommodate digital information was accompanied by an exponential upward tendency of new data colliding in the 80s with the mounting constraints regarding data warehousing infrastructures. Furthermore, this scenario is emphasized by the manifest limited capacity of hardware equipment at the time (Hilbert and López, 2011). Hence, emerged in the 90s the advent of Big Data epitomized as the *data boom*, defining the amount of data greater than the capacity of the regular computer. Thus, one could neither be processed/stored by regular hardware; nor the analytics performed by standard software, regardless of its structure (Cox and Ellsworth, 1997; Mayer-Schönberger and Cukier, 2013).

Beyond the original virtue (of solving data-related constraints), Big Data opened a myriad of opportunities for information building. With the rapid development of a triad of (*advanced; exploratory* and, *discovery*) *Big Data Analytics* techniques, a broad instrumentalization potential has emerged, from scientific investigation to market research, in a wide variety of domains such as medicine, economics, mathematics, law, philosophy or business; therefore, attracting a considerable attention of the academia (Russom, 2011). Fueling the phenomenon of BD Analytics´ diffusion was, among others, the proliferation of sensor-enabled and internet-connected devices digital communication, across smartphones, household appliances, vehicles, industrial machinery, equipment and even humans, also referenced to as the Internet of Things (IoT). The IT consultancy firm Gartner (2013) assessed that by 2020 there will be 26 billion of such devices in the world corresponding to averagely 3 devices per person in the world.

In this context, its utilization with a market orientation came naturally, both through an *outside-in* and *inside-out approach*, as described below. Uncovering associative patterns between units of information undetectable by conventional means, and consequently, allowing the creation of meaning systems of rather utility, BG resembled in the first decade of the 21st century the "holy grail" of operational advantage building for re-strategizing business competition (Lambrecht and Tucker, 2015). Furthermore, it triggered the fashioning of modern business *analytics strategies* for maximizing market intelligence. Firstly, the companies were able to remedy their inner (multi-tier) vulnerabilities enterprise-wide, and simultaneously leverage their system´s operational efficiency (outside-in approach). Secondly, it democratized the use of data-driven insights for the gain of competitive parities or firm-specific advantages (FSA) (inside-out approach). An example of its applicability is the use of BD on predictive maintenance of an automobile, as sensors collect data records on multiple parameters (e.g. temperature, vibrations, oil pressure, idle time or fuel consumption) consequently preventing eventual breakdowns before they actually occur. Other benefits within e-health networks, accounted the mining and sharing of patient x-rays, CT scans, and MRIs with analytics levering cost-reduction and accuracy of treatments (Michael and Miller, 2013).

Such technological big-bangs had then the ability of shaping social interaction and convert it into valuable data for several quadrants of society as communication, security or consumed packed goods (CPG) whether provided by private-equity firms or public institutions (Abrantes, 2020; Van Dijck, 2014). In this context, the way to approach analysis and projections has naturally changed, with the consequent devaluing of sampling tools and its precision being no longer an imperative, as Big Data dealt almost entirely with unstructured and heterogeneous units, and the sheer volume of data compensated data inaccuracies or "messy" data (Gandomi and Haider, 2015). Therefore, the virtues of Big Data analytics derived from a shift from human behavioral´s causality into an interpreting of the *factum,* the unravelling of patterns for the estimation of future events, foreseeing unpredictability, forecasting scenarios and eliminating human error and bias from organizational processes (Baruh and Popescu, 2017). Unsurprisingly, the virtues of BD spawned a wave of euphoria labelled as *dataism,* intrinsically assimilating the level of trust assigned to businesses and public institutions pertaining to private data and the faith on harmless use. Within the myriad of exchangeable units of petabytes, it is comprised the personal data created by customers while interacting with media channels (i.e. *digital footprints)* (Muhammad, Dey and Weerakkody, 2018).

However, 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 unstructured and heterogeneous personal data, as life events or discrete social interactions, as valuable and monetized insights, and therefore, discernible as potentially hazardous and noxious for the individual, here identified as the personal *data-owner* (Van Dijck, 2014; Gandomi and Haider, 2015). With this regard, Zuboff (2019) dubbed the prior controversy as a totalitarian pattern of a surveillance capitalism (*dataveillance*) imprisoning citizens and subordinating those to behavioral modifications, as predictable events with an exchangeable human stock actionable value, tradeable as units of behavioral data. In this context, *data-brokers,* as the entities responsible for the direct and/or indirect manipulation or final-utilization of those stocks of data are perceived as hazardous entities with a questionable moral conduct. This occurs, not solely on the grounds of the gained ability of the firm, to anticipate future human actions with BD; but foremost, because of the asserted anti-democratic self-empowerment of the firm, to condition and alter what an individual is able to perceive, think, see or feel, as a "causation process" modifying in a deceptive way, among others, the patterns of choices and consumption.

Consequently, as Big Data is gradually pushing the scientific orientation from causation to correlation, conveying the hazard of results biasing on future statistic projections and the risk of the *apophenia* comprised in the overanalyzing or finding false positives/false negatives (Ahn and Lee, 2020; Dekimpe, 2020*).* For instance, Leinweber (2007) statistical association between the annual changes in the S&P 500 stock index and butter production in Bangladesh exhibited a strong relation, yet a spurious one. Mayer-Schönberger and Cukier (2013 argue that the root of apophenia lies on a deliberate intention innate to behavioral economists to foresee causes even when none exists, emphasizing how critical it is to avoid own cognitive biases from deluding us, and simply let the data speak.

The latter constituted the initial problematization for a pursuing such a research avenue on *digital footprints*. Here, we account both natural data from pure digital footprints and *data fumes* as the subset of digital footprints descendant from its prior application and contrived datasets. Hence, on the epicenter of such wrangling of the parts, we have set in forehand an axiological assumption of equidistance to the problem stated and to the apparent conflictuality between the individuals´ data ownership (data-brokers) and the organization´s utilizing of their data, the data-brokers.

1.2 Gap-scoping and research aims/objectives

As an emergent field, BD earnt an exponential audience of academics and rise on publications in recent years with a significant coverage of the domain of digital footprints and sentiment analysis. Nevertheless, the studies on the Nordics cross-observing *rationem* (reason) and (*motus*) emotions regarding the utilization of personal data, observing the intertwining of individual prior knowledge (*digital footprint awareness*) and his/her acceptance (*digital footprint sentiment*) are still on an early-stage specially in the context of Danish market. Therefore, it ought to be clarified, the singularity our research resides first on crossed investigation of these two underlying components of individual´s digital data, but also on the fairly untapped incumbent´s side, particularly in the Danish tertiary sector, the latter justifying the choice of the marketplace of observation.

Hence, our purpose here unveiled is to grasp whether the individuals acknowledge: (i) the utilization of their personal data by third parties; (ii) their endeavors to prevent or interrupt it (awareness); and, (iii) their acceptance of the conditioning of choices due to BD analytics (sentiment). Likewise, our enquiry mirrors the same components on the data-broker´s side to understand: (i) the perception harm of the brokers´ own activities; (ii) feedback on consentment and satisfaction from the counterpart. Thus, the general aim (A) of this research is to understand and explore the perceptions, sentiments and behaviour of consumers and data brokers in relation to their mutual data exchange. Therefore, a set of objectives derived from the above:

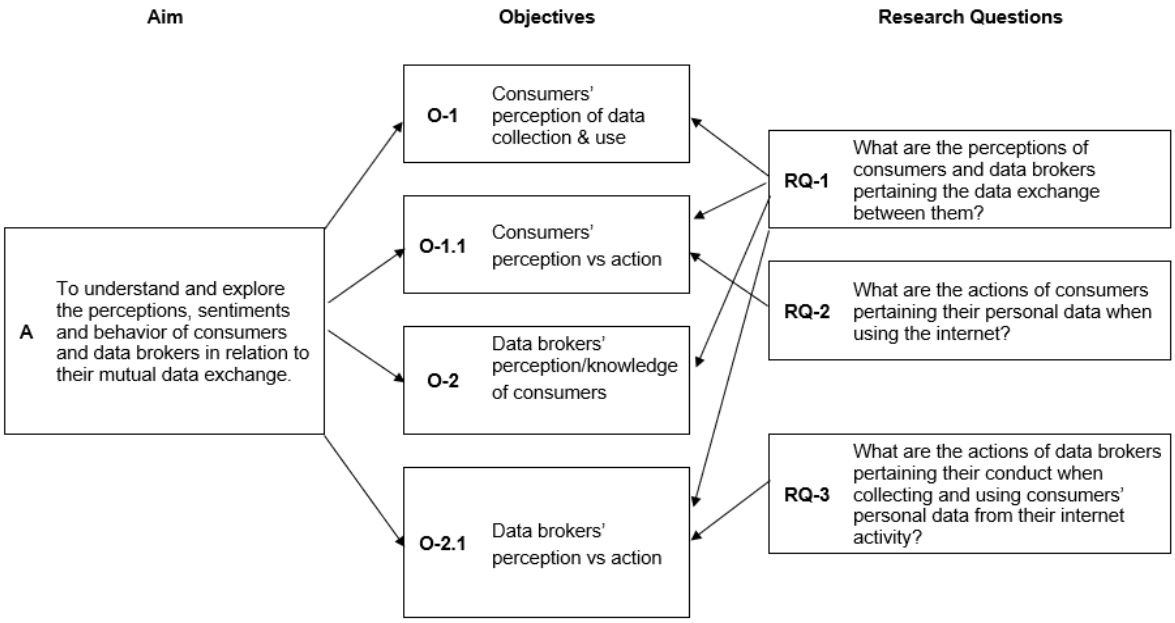
Objective 1 (O1 - Data-owners' perceptions): To comprehend the consumer´s general perception of data broker´s conduct processing their personal data.

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

Figure 1 - Research Framework



Source: Own elaboration

The figure represents the association of aims, objectives and the formulation of research questions for testing the above, with a general aim deriving into four 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 are simultaneously here enquired, through internet-mediated interviews and questionnaires with the outside-research team´s assistance. Accordingly, the general structure of this manuscript proceeds with a theoretical revision of 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. Then, it proceeds with the description of the research design, describing the typology of the case, the demographic profiling of the partaking firms and sampled units of analysis, both participants and respondents, and with a clarification of the rationale underlying to the choice of methods and its collection momenta. Subsequently, the research continues with 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 relation to the aim and objectives delivered in the research framework, complemented by the managerial implications and the opened avenues for further exploring and testing this topic and subsequent knowledge-building.

2. Literature Review

Big Data is a phenomenon of increasing ubiquity and henceforth, literature associated with the phenomenon is becoming inherently copious (De Mauro, Greco, and Grimaldi, 2015). The vastness of the theme led these researchers to immerse into a theoretical revision associated of the problematic addressed in the previous section, as to the sentiment of digital footprint´s abusive exploitation. The definition here adopted of Big Data is De Mauro, Greco and Grimaldi´s (2015) attempt of a consensus, condensing parameters observed on other definitions:

> "*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*." (p. 103)

The definition is parameterized to cover the most components addressed in other definitions while remaining concise to distinct the phenomenon from other boundary conceptions of large datasets. Such conception of Big Data pertains to minimum one of the following themes:

1. Information, the fuel of Big Data: From digitization of analog information over mass digitization to datafication (organizing and cross-referencing digital information to generate insights).
2. Technology: With an exponential increase of data availability with proliferation of Internet of Things (IoT) comes a requirement of an equally exponential increase in data storage and data processing capabilities.
3. Methods: The bottleneck of Big Data exploitation perhaps presently lies within the Method, or the actual analysis of the Big Data. The human element, or the "algorithmists" (Mayer-Schönberger and Cukier, 2013), are not hatched at an equally exponential rate as the data, which they must analyze.
4. Impact: The actual and potential impacts of Big Data across multiple areas of industry and science are indisputable. Big Data is becoming a means to competitive advantage, and companies are challenged with keeping up, both in terms of implementing Big Data technologies, acquiring competencies and acquiring the data, which it is all about.

The digital age leveraged by the Big Data´s revolution had the virtue of breaking strings with the dominance of analogue media and disrupt data storage capacity (Hilbert and López, 2011). On the 50s (20th century), data was considered "big", when approaching the 5 megabytes mark, corresponding to the hard drive (HD) capacity of IBM's RAMAC 350 (Hoagland, 2003). On the 80s, digital data corresponded solely to 1% of data archives (Hilbert and López, 2011). With the turn to the new millennium digital data storage weighted equally as analogue media (50/50), and in 2007, 94% of all data was already stored digitally (Hilbert and López, 2011).

The advancements in ICT hasted by Big Data ("*data whose size forces us to look beyond the tried and true methods that are prevalent at that time*"), brought a societal´ s three-dimensional perspective towards data utilization (dubbed as the "3Vs" – volume, velocity and variety), which opened new horizons for a data-driven second economy (Jacobs, 2009, p.

44). The drastic expansion of the order of magnitude of data which one could manipulate (Figure 2), allowed the scalability (*volume*), increase of pace (*velocity*) and the accounting of a broad span of information assets (*variety*), acknowledged by multiple scholars as key economic assets for the future (Cavanillas, Curry and Wahlster, 2016; Kitchin and McArdle, 2016).

Figure 2 – *Data Storage: Standard units of measurement*

Unit	Value	Size
bit (b)	0 or 1	1/8 of a byte
byte (B)	8 bits	1 byte
kilobyte (KB)	10001 bytes	1,000 bytes
megabyte (MB)	10002 bytes	1,000,000 bytes
gigabyte (GB)	10003 bytes	1,000,000,000 bytes
terabyte (TB)	10004 bytes	1,000,000,000,000 bytes
petabyte (PB)	10005 bytes	1,000,000,000,000,000 bytes
exabyte (EB)	10006 bytes	1,000,000,000,000,000,000 bytes
zettabyte (ZB)	10007 bytes	1,000,000,000,000,000,000,000 bytes
yottabyte (YB)	10008 bytes	1,000,000,000,000,000,000,000,000 bytes

Source: TechTerms, 2012

Yet, in the current "petabyte age", the exponential growth or "revolution" inflicted by BD was accompanied by an equivalent rise in its market value (Lerman, 2013; Manovich, 2011). The empirical research of Kitchin and McArdele (2016) over 26 datasets refined the threefold logic (3Vs) deepening the understanding of the BD to *seven traits*, adding the properties of *exhaustivity*; *resolution* and *indexicality*; *relationality*; *extensionality* and *scalability*.

These datasets with enormous quantities of data (large volume); generated on a continuous way, handled on an on-going/real-time basis (velocity); and, accommodating variety (structured/unstructured and semi-structured), observed also properties of *exhaustivity,* as the datasets evolved from samples of small (typical survey or administrative samples) to entire populations capturing entire systems of behavior (Mayer-Schonberger and Cukier, 2013). Furthermore, they accounted 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*, as it allowed the incorporation of new fields and their scalability as to their expansion in size (Marz and Warren, 2012).

Hence, the morphology of BD ascribed to be undeniably associated with *value,* due to the insightful depth and great utility extractable there from, positioned the data-brokers as the "wizards of mining" envisioning their great potential as *value creators* (Marr, 2014). In this context has emerged the *Big Data Public Private Forum (BIG),* also designated as the *"Big Project",* a public-private effort of 11 European entities. A set of three entrepreneurial universities have adhered to the project (the *National University of Ireland;* the *University of Innsbruck;* and, the *University of Leipzig*). Six other entities, from civil society have participated, including industry practitioners (*AGT Group R&D GmbH; Atos Spain S.A.*; *DFKI - Deutsches Forschungsinstitut für Künstliche Intelligenz; Exalead; Press Association; Siemens AG; STI International; The Open Knowledge Foundation Deutschland)* have cross-collaborated with the first ones; in order to, explore the BD technologies, and furthermore, support the European Union´s (EU) establishment of a research roadmap for Horizon 2020, as to the implementation of an agenda for the Big Data economy (Big Project, online).

The *Big Data Value Association (BDVA*) was then established in 2014 and furthermore mapped the BD´s lifecycle and the *Big Data Value Chain (BDVC),* with the latter comprising the components of each phase of the BDVC (1. *Data Acquisition*; 2. *Data Analysis*; 3. *Data Curation*; 4. *Data Storage*; and, 5. *Data Usage*) (Cavanillas, Curry and Wahlster, 2016). Furthermore, it emerged the *Big Data Value contractual Public Private Partnership* (BDV-cPPP) as a framework for guiding BD research and industrial practices towards economic systems based on the assumption of BD´s cross-sectorial requirements and political, economic, social and technological (PEST) challenges for data-driven industrial leadership, investment, and commitment.

2.1. Digital footprints: wrangling and challenges

To the arbitrate of 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, as specific regulation to the protection of unrestricted movement of such data. This legal act was designated as the *General Data Protection Regulation* (GDPR) – *Regulation (EU) 2016/679*

(Eur-Lex, 2016), including among others regulation 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: (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 resemblances acknowledged as to the prior Data Protection Act 1998 (*1998 Act*), the GDPR neglected a principle for the protection of individual rights and for the international transference of personal data (ICO, 2018).

Such omission is attempted to be (partially) 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 law enforcement is growing in EU, especially as to the commercial usage of personal data by corporations (as presented at the introduction section). Such friction seem to be stilted by a confrontational scenario, fueled by the incumbent's practices of *hyper-targeting* of consumers, in confrontation with, privacy rights (including individual ones) observed in the acts above. Furthermore, the GDPR touted on *article 17* the "*right of erasure"* or the *"right to be forgotten* (RTBF)" below described. Likewise*, article 37* defined the creation of a mandatory role, the *Data Protection Officer (DPO)* for all organization 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 organisation [Large and small and medium-sized enterprises (SME)] that create, collect, aggregate, and transform data from both public and private sources"* (Cavanillas, Curry and Wahlster, 2016, p. 52). The second (technology providers) as "*typically organisations (Large and SME) as providers of tools, platforms, services, and know-how for data management."* (Cavanillas, Curry and Wahlster, 2016, p. 52).

Such *hyper-targeting* (also designated as *micro-targeting)* gives attention to an exhaustive utilization of detailed customer-data and marketing automation tools, based on predictive modelling, 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).

Hence, the processing of such informational assets towards value-creation relies typically occurs on a *three-stage model* of automated data processing. Firstly, data is retrieved from multiple data sources to a scalable storage solution such as a *NoSQL* database or the *Hadoop Distributed File System (HDFS)*, then reorganized and ordered and purposively crunched on locked-in or open source protocoled big data software using by analytics algorithms (e.g. *Casandra, Cloudera, Hortonworks*, or *MapR).* The latter aims at setting promotional campaigns designed to appeal to micro-market segment groups of potential customers (Kim, Trimi and Chung, 2014; Cavanillas, Curry and Wahlster, 2016).

Yet, if controversy towards BD focusses mainly on digital footprints, defined in the first section (i.e. customer´s personal data generated in interaction with media channels), however, evidence reveal an exploitation of data with both commercial and non-commercial purposes (Muhammad, Dey and Weerakkody, 2018). We recall here one of the most striking examples of the latter, the 2015´s scandal of *Cambridge Analytica* on the North-American presidential campaign (González, 2017). Nonetheless, scholars raise doubts as to the absolute effectiveness of BD, as some studies BD´s advertising performance exhibit contrary results. Here, Semerádová and Weinlich´s (2019) research over 840 Facebook ads, concluded in fact a limited effect and counterproductive results in terms of user reactiveness, suggesting a negative correlation of overexposure to hyper-targeting and future purchasing causality, as the first may lead to negative attitudes toward the advertiser.

Yet, such negative sentiment finds real ground implications on privacy loss, self-determination and on price discrimination and reduced consumer welfare. Hence, some scholars have postulated the importance of "*stopping the unjustified accumulation and commercialization of personal data*" and furthermore advocating an upper responsibility of

consumer protection, which ought to be vested on the legislator hands to immediate regulate and execute the compliance of practitioners to it (Lecuona an Villalobos-Quesada, 2018, p. 291). This comes in a current scenario where further concerns were raised towards the public sectors' equivalent exploitation of Big Data envisioned as a George Orwell-ish "1984" dystopia.

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 hazard 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 malefic instrumentalization of artificial intelligence (AI) for a symbolic manipulation of the masses. Conversely, consumers are perceived furthermore as unprepared to understand or evaluate the ethical implications and the knock-on effects on their lives of usage of their digital footprints and subsequent limitations forced to their actions, underlying to the groundbreaking and transformational nature and pace of Big Data´s development (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 it has been associated with a deceptive molding of consumer's thoughts, and their intentional redefinition of behaviors (causality reasoning). Therefore, some scholars have alerted to a form of digital captivity or servitude, as a current state of Big Brother´s social regime, of post-modern digital voyeuristic unfreedom and surveillance capitalism of (quasi) totalitarianism (Zuboff, 2019; Diamond, 2019; Brayne, 2017; Xue et al., 2016)

However, if the prior, surrounding the Big Data utilization´s ethics, especially its commercial, symbolic and political manipulation and surveillance of masses collides with the principles of utilitarian business ethics worldwide, within the EU´s sphere such problematization gains wider contours. Herein, these perceived (un)ethical practices clash towards a primary ethical principle of legal ethics (Nielsen and Andersen, 2008); where ethics defines an attribute to define a decision that is both legal and morally acceptable, and unethical "is either illegal or morally unacceptable to the larger community" (Jones, 1991, p. 367). 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 leveraged then a worldwide debate as to a superlative right of the individual privacy (Xue et al., 2016). Here, the RTBF portrays the right to a digital oblivion of incidents or events (*footprints*), being the "controller" (data broker) obliged to personal data erasure without undue delay, as the latter retains no overriding legitimacy to further process of data, being such action considered unlawfully processing.

Nevertheless, business-related practices exhibit a partial compliance to the GDPR, deviating from legal principles. For instance, RTBF requesters of delisting or erasure (e.g. from a service newsletter), face the unprotected assault of data reutilization (*Streisand effect*) and its reconditioning (*data fumes)* (Xue et al., 2016; Thatcher, 2014). The first effect encompasses the triadic rights of whether to hide data, to remove it and/or to censor a piece of information, since the unregulated and uninspected practices of third parties republishing of the original (made-public) data remain inextinguishable (Xue et al., 2016; Thatcher, 2014). Although, if this effect refers to natural data originated from digital "base behaviors" and its inherent threat of exposure due to reutilization, the data fumes correspond to a descendent exhaustive extraction procedure of third-party vendor, and the dangers of biases from capitalization from contrived datasets, as the "fumes" of perceived behaviors and not their manifestation (Thatcher, 2014).

Consequently, a mix of disappoint and skepticism surrounds the GDPR deriving from the EU State-members and EU´s institutions inability to ensure a full compliance of principles and dispositions, namely the articles 37 and 17. The latter issue, as to the a RTBF effectiveness has eroded as to the article 17 forces the controlee to take action, which has not effect upon the Streisand effect and data fumes steadily growing both phenomena. Therefore, this regulation epithetized with lack of consistency, as it stimulates data-broker´s prevarication and non-compliance, and the inaction of consumers is penalized. Thus, the GDPR is at some extent professed as a modest expansion of data privacy rights. This problematic 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.

2.2. Digital footprint´s awareness

2.2.1. 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 of 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, Chiang and Storey, 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 thoroughly appreciated by the masses.

On the other hand, data-driven marketing backfires through fixating consumes 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 a loss of privacy and furthermore the normalization of privacy loss, exemplified by commonly trivialized practices of 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 lack of individual capabilities. 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 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, an upper threat is the *First-Degree Price Discrimination* aiming at individual consumer's absolute maximum price or reservation price. Typically, this is theoretically possible only in a monopolist market structures, but Steinberg (2020) asserts that advancements in Big Data identifying individual consumer's reservation prices may soon be possible even in a perfect competition scenario using Big Data. Yet, Mayer-Schönberger and Cukier (2013) asserts a more likely scenario of analytics exacerbation within central planning economic systems, where predictive policing, identical to consumer´s behavior prediction; however, with a broader societal radius it is applicable to law enforcement and intelligence communities. Spatial BD is then utilized to correlate date, time, type, and location of events, as car accidents or crimes), and accounting historical data to identify "hot spots" as future focus areas. An example of the prior within a free-market/private-enterprise system is accounted by Crawford and Schultz (2014) in the United States of America (USA) by Maryland´s state police surveillance of human rights groups, peace activists, and death penalty opponents over a nineteen-month period.

2.2.2. 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 simply understanding customer behavior, to a better segmentation and targeting, until the 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 is pushing 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 anticipates a bright future plank for a newer "digisense" era of abundant interrelatedness of sensors, controllers, big data and data science (Gartner, 2013).

Consequently, the market size of this technologies is expected to reach by 2022 the USD 6 trillion USD mark (IDC, 2018). For instance, Walmart, the number one retailer in USA, which pioneered on BD analytics on its industry, 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 utilization 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, Lahcen, and Belfkih, 2015; Marr, 2017).

On the public sector, the traction gained by Big Data favoured a variety of applications ranging from education or health, increasing citizen's engagement in public affairs, prevention fraud/crime, improving national security, and supporting other forms of well-being (Kim, Trimi and Chung, 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 past track of building fires to establish fire risk predictors, anticipating incidents and optimizing the daily inspection work (Mayer-Schönberger, 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, raising the legitimacy and ethical issues as to its adoption and uncovering the debilities of the GDPR and insufficiency of regulation (Xue et al., 2016; De Mauro, Greco, and Grimaldi, 2015; Thatcher, 2014).

3. Methodology

The research framework diagrammatically represented in section *1.2 Gap-scoping and research aims/objectives* aggregates the tenets of the theoretical testing here followed, as the researcher team intended to gain an insight over the extent of conformity between sentiment and behavior towards digital footprints. Moreover, it aims at unveil whether consumers are willing to prescind of consecrated privileges or rights on daily actions and comprehend the trade-off (benefit-cost) and its baseline of marginal benefits/costs equilibrium.

A post-positivistic stance as the research paradigm adopted, combines a one-phase mix-methods applied in a concurrent manner. The prior derived from a seminal reflection on a dyad of considerations. Firstly, the aims/objectives and research question drawn as the research angle to be pursued. Secondly, the assumptions raised as to the *locus* (problematization) and underlying gap, cognitive attributions (including limitations) of the research team and the ethical bond/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.

Each case-firm had one qualitative unit of analysis (UA) representing the data-brokers, and multiple quantitative ones obtained from their customer-base, being therefore an identical analytical procedure between cases, assuming naturally an iterative logic with embedded UAs. In total, were collected 2UAs from the case-firms using open-end interviews (I_n) and 137 UAs gathered from closed-end questionnaires (Q_n) from data-owners. The profiling of both case-firms and the participants/respondents is presented in Figure 3 and 4. 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 quants respondents (data-owners) without the case-firm´s intervention were targeted through a social media platform, using personalized messages. As to the latter, the response rate achieved a high mark (.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

* expressed in Thousands of units (K) –
Includes permanent/temporary employees (year/$\bar{x}$)
** Billions *Danish Kroner* (DKK)

Source: Own elaboration

Table 2 - *Participants (Pn) and Respondents (Rn)*

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[1]	HC[2]	-

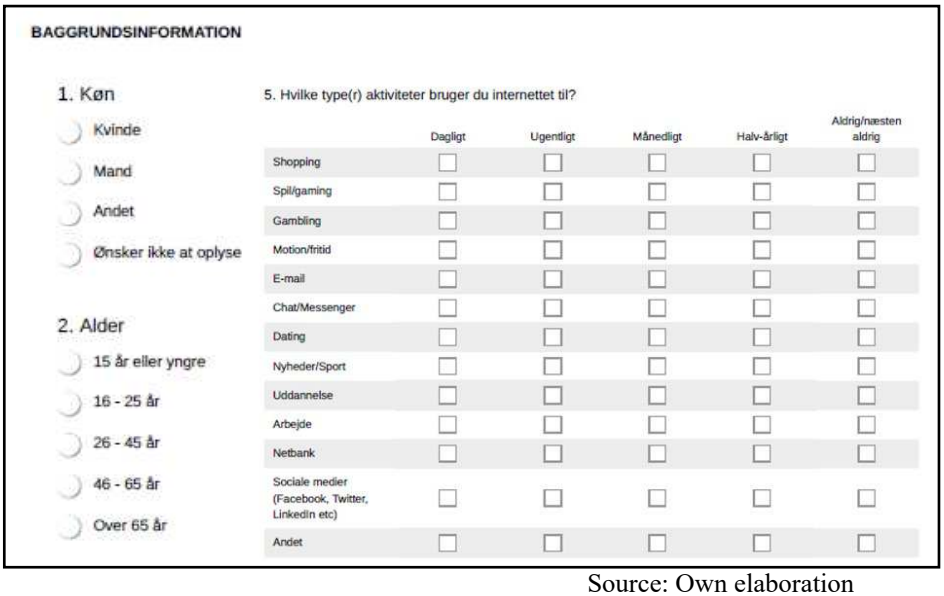
* Levels in accordance to the European Qualification Framework (EQF);
[1] Vice-President, Business and Portfolio Planning [2] Head of Controlling & Real Estate
Source: Own elaboration

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, were instrumentalized opinion and behavioral variables, 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.

Exhibit 1 – *Excerpt of method 2 (quants - questionnaire)*


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	☐	☐	☐	☐	☐

Source: Own elaboration

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 section *4. Data analysis and findings.*

4. Data Analysis and discussion

4.1. 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:

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

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*

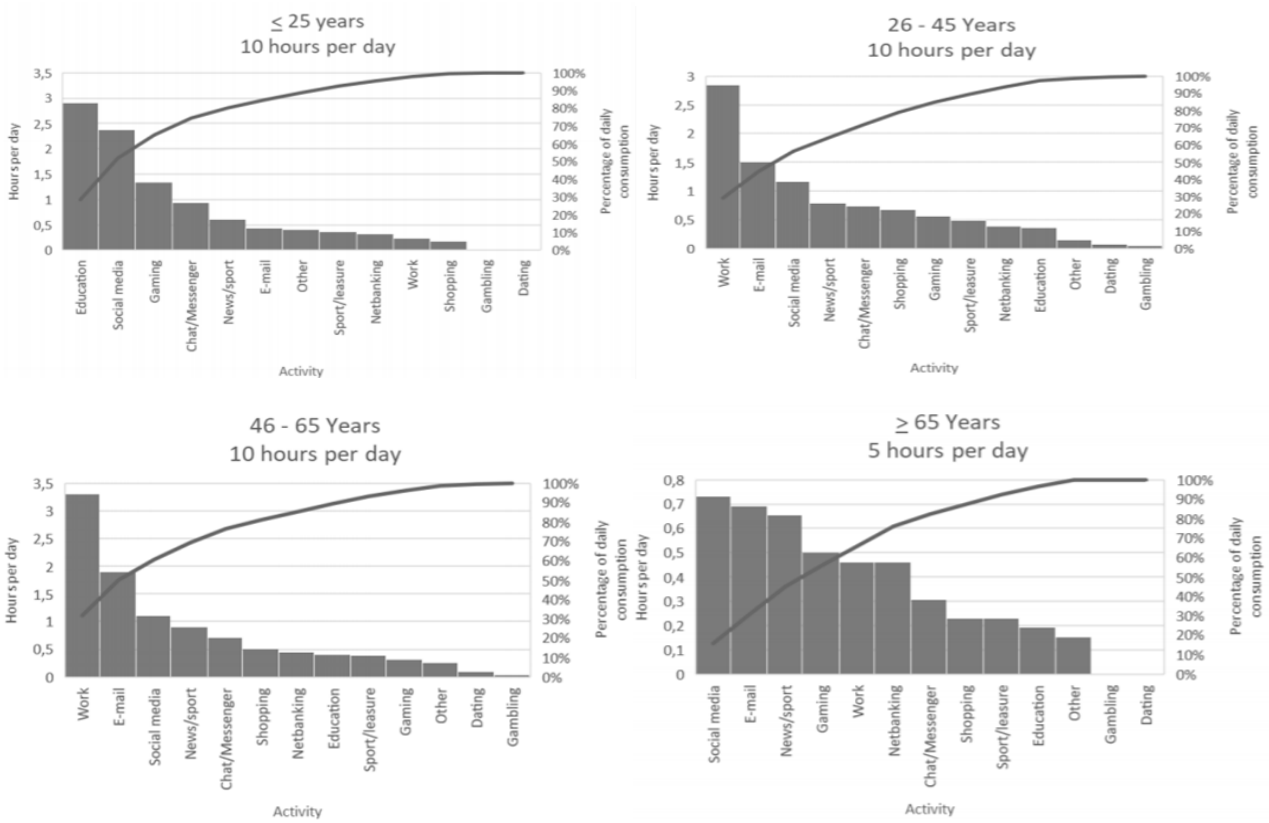
Source: Own elaboration

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?").*

The interpretation of the results from interval variables of the respondents' answers to IQs using non-parametric testing instrumentalized the correlation coefficients, as to Evans (1996) levels of significance. The latter varying from a very strong negative (-1 and -.8) to very weak negative (-.19 and -.01) correlation; and from very weak positive (0 to .19), moderate (.4 to .59), to very strong positive (.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 in order 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 - *Sum of Internet activity per class (age group)*



Source: Own elaboration

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

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	.07	.01	.74	.06	.1	.04	.06
B. Changing cookie settings	.16	.23	.64	.06	.1	.1	.09
C. Opt out (privacy policy unacceptance)	-	-.02	.64	.07	.1	.13	.06
D. Deleting browsing history	.13	-.01	.4	.18	.21	.13	.1
E. Using incognito-mode	.27	.27	.58	.06	.07	.15	.14
F. Blocking website	.12	.08	.63	.12	.14	.06	.05
G. Manually scanning for malware	.11	-.03	.61	.14	.18	.07	.01
H. Automatically scanning for malware	.19	-.09	.18	.05	.09	.19	.49

Source: Own elaboration

The majority acknowledged never (or almost never) have applied such protection measures. The least 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 measure most frequently and actively applied is the deletion of browsing history, which is done at least monthly by 43% of the respondents. The protective measure most frequently, but passively applied, is the scanning for viruses using 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 (.11 to .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 (.23 and .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).

4.2. (Data-brokers´ qualitative) manifest content and Thematic Analysis (TA)

As to the comprehension of the data-broker 'side endeavors, a thematic analysis (TA) method is applied, as to the exercise of fashioning discrete theoretical themes (codification design) and subsequently consecution of a *Gestalt* analysis for unravelling 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, 2015). Hence, our instrumentalization of TA analysis is rather centered on *primary saliency,* as the linkage of launched signifiers and into realized signifieds (Buetow, 2015).

An open coding procedure aiming at illustrate each theme, instrumentalized 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 manifest 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 - *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/wellbeing	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

** CId – Code´s identification; ** sentiment equaled to "perception" on the research framework*

Source: Own elaboration

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 data-owners delivered value. The prior 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 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.

4.2.1. 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 on 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 (behaviour) 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 (.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, advocating 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 whom 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 section 2 (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 sub-themes 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 behaviour 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.

Table 6 – *Qual Data-theory (BDAR) triangulation*

Theme	Sub-theme	Data (Qual)		Sub-theme	Theory (Section 2)	
	(ST_n)	Sub-theme	Implication	(ST_n)	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/counselling	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/wellbeing	Data-brokers Data-owners	T38O	Economic and business development (2^{nd} economy; 4^{th} 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

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´ (BDAR) resources for commercial purposes. From the absolute frequency, 3 of them co-occur (*f=.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, have incidence on data-owners. 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 (.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 representant) did not generate any quotations, as perceived evidence of this phenomenon.

As to the perception of the future of these digital footprints with regard to opportunities (*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 implication for the future, here affecting dominantly the data-owner´s side *(.8333)*. Furthermore, these exhibits treated on the logic of sense giving to understand the *foci* of the social fabric as to the perceptions of digital footprints.

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				
	T49T		T4-T				
	T21W	Data Fumes	T2-W				
	T43T	Data pollution	T4-T	-	2	3	3
	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

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 influence of the BDAR and the digital footprints, accounting for .225 and .2 respectively. Furthermore, considering these outputs, we have modelled the *impaction index* (II) of BDAR and digital footprints, considering the relative impact (RI) of each theme (T1 to T4).

$$II_i = Sum\ (current\ positive\ implications\ /\ overall\ positive\ implication) + \Delta\ (T_n) \qquad (1)$$

$$II_i = (RI_{(data\text{-}brokers)}\ U\ RI_{data\text{-}owners}) + \Delta_{t=1}\ (RI_{data\text{-}brokers}\ U\ RI_{data\text{-}owners}) \qquad (2)$$

Considering furthermore,

$$RI_{i\ (data\text{-}brokers)} = RI_{i\ (data\text{-}owners)}$$

Where as

$$RI_{i\ (data\text{-}owners)} = \sum\left(\left(\frac{S}{OS} - \frac{W}{WT}\right) + \left(\frac{S}{OS} - \frac{W}{WT}\right)\right) \qquad (3)$$

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.

$$f(x) = SWOT_{(t1)} - SWOT_{(t0)} \qquad (4)$$

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 *1,57143)* denoted a better exploitation function of BD for the firm´s side than the individuals.

4.3. Discussion

By the empirical testing disposed on the section 4, we observe the great benefit´s gained by the data owners on instrumentalization of their digital footprints as BDAR for commercial exploitation came at the expense of legal and economic consequences for the consumer-side. On the legal landscape, this encompassed the unprotecting of legal rights (namely the ones consecrated on the GDPR described in section 2) 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, and we argue to be a consequence of the first, 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 BDARs, modifying and/or extending portfolios and refining digital positions, and oddly, the interviews did not pinpoint any drawbacks for them or the consumer regarding Big Data. 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 a fake data fume for consumers´ manipulation and polarization of products/services on large and dominant incumbents. This a phenomenon which one participant pinpointed a belonging to a larger one, the growing data pollution on multiple societal quadrants and with multiple purposes and unpredictably spiraling up. Here, one of the participants stressed how unprepared organizations

are nowadays to deal with informal data breach, which may even complexify even more the current typology, flow and directionality of BD in circulation. Yet, also opportunities to data-brokers were 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 the immersion of a second economy and a new (4^{th}) industrial revolution with inherent benefits for broad scientific developments as well.

The perceptions and actions revealed though a clear gap accentuated on more 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 predominantly negative opinions towards Big Data amongst consumers appears disproportionate to the apparent limited application of protective measures by the same consumers to mitigate possible undesired effects of dataveillance and analytics practices, denoting 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 consumers seem to rely faithfully on GDPR protection rather than self-protection, a phenomenon acknowledged also before by Baruh and Popescu (2017). As to altering of the prior, women respondents exhibit a higher pre-disposition for disbursing 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 applied to the Nordics applicable to the Danish market and furthermore pursue new avenues for enlightening the extent of influence per theme and compare them through replication on multiple research angles.

5. Conclusions

The initial glowing phase (dataism) has clearly vanished. Two decades ago, the marvel of the digital markets and its inherent economic benefits seemed to attract steeply both the demand and supply sides towards platform utilization, consumption and competition. However, these days have dissipated with the commoditization of those benefits through a continued utilization, furthermore emphasized by a marked accentuation of the exploitation models of consumer´s digital footprints. As referred by one of the participants it is a "…arm race…". Big Data is a "need" and companies identified an element of "stick" behind this necessity.

Yet, the growing discontentment and contesting of the effectiveness of personal data protection and the inoperance of individual privacy laws (blistered in EU by the GDPR ineffectiveness as discussed in section 2) is nowadays exploited by abusive, deceptive and/or illegal practices of several firms (e.g. hypertargeting; preferences´ manipulation; price steering; or, price discrimination). Such scenario has steered into a conflict between parts. The prior is furthermore emphasized by evidence of data breach, and dataveillance connoted the latter as being as totalitarian voyeurism of corporate practice forcing and slurping the consumer to the maximum of its the capital resources, using sophisticated BDAR with algorithms fashioned to explored beyond data fumes and/or generate data pollution or fake data.

From this empirical testing, it is clear that a tripartite collaboration is essential (governmental authorities, firms and consumers) considering the consumer´s inability towards self-protection and safeguard of his/her interests and inefficient regulation. The perception of personal data utilization (*objective 1 – O.1*) as described in the previous paragraph, is not accompanied by according actions (O1.1.) prevailing a hazard *laisser faire* on the Danish market. We argue, social regimes ought to alleviate the burden of individual responsibility of the consumer own surveillance of the compliance of the organizations with their personal data and privacy regulations. Thus, we argue that on the national/meta-national authorities´ side, there is an imperative to introduce further mechanisms of collective protection and safeguard of citizen´s rights. The path advocated here is, stronger regulation, stronger inspection mechanisms to minimize the horizon of deception and/or prevarication or theft, and underlying law enforcement verifying their applicability. Here, we corroborate with Zwitter´s (2014) postulation of the re-conceptualizing of the privacy, personal data ownership, data fumes, digital-crime, guilt, and the circumscription of the scope of crime and crime prevention. Here, we do argue towards global approaches for legal bids and towards a homogenization of the regulation of personal data and privacy.

Regarding the gap (perception-action) the bottom-line issue here drawn is the forcing of the consumer to act to ensure the effective observation of an individual right, which seems to lay on the "against will" or undesired action (*contra naturam)*

and the underlying (un)awareness for the consequence of inaction as personal data protection. This is equivalent of criminal law to determine weapon possession as mandatory to the self-prevention of getting car hijacked or physically assaulted on the street. Thus, there´s definitely a road to be travelled by policy-makers in Denmark and EU on this matter. It is advocated here an intertwining of the prior with the upskilling of consumers through the mediated hand of public authorities´ stimuli to further capabilisation on data protection.

As to the data-brokers´ side *(objectives 2 and 2.1.- O2; O2.1.)* companies face a reputational challenge, on the current climate of dataveillance, being on the consumer´s eyes the offender or the prevaricating part, bending the GDPR boundaries, and so invading lives and usurping financial resources and wealth. At firm-level´s intervention, data revealed an equivalent gap of perception of pros/cons for brokers and owners versus action taken (O2.1), as the companies observed a comfortable isomorphism favorable to business. This reversion of reputational and economic risks requires the adoption of a triadic approach (ethical, technological and socio-political). Beforehand, companies hold a legal responsibility of complying with legal rights scrupulously as to the GDPR dispositions, and on the virtue ethics domain, here they may add the good will and responsibility of balancing the distribution of power as to the effective exploitation of the administered data with their consumers. This implies three foundational actions. Firstly, tracking the origin of data banks before its utilization and spotting traces of pollution (including contrived forged or fictional data). Secondly, constructing and managing the consumer´s digital footprint path for traceability purposes and reversibility of displays. Thirdly, incorporate shielding technology and data governance procedures as to data warehousing (and shared-warehousing) infrastructures for ensuring watertight access, avoiding leaks (data breach) whether disrupting inside-out/outside-in holes and breakouts. Here, the researchers argue on favor of a superlative model beyond the synchronization of access, as an orchestration of real-time joint-administration of personal data with facilitated self-administrated access to the consumer, with transparent traceability onto the far-end dissemination of data and self-erasure configurations. This model ought to bring together and unify the three main agents (owners, brokers and authorities). Yet, at firm-level, data-brokers endeavors are recommended to be enlarged through corporate responsibility policies and social goals, in parallel to the above, applicable to the whole value systems (including suppliers, distributors/retailers and licensees) with the purpose of monitoring/controlling own governance practices and conveying transparent results to third parties, as intelligence screening instruments as *Sustainalytics* or other equivalent apparatus. Finally, a model with this fashioning discards practices of algorithmic refinement of predictive models from data fumes, assuming furthermore an upstream responsibility for tracing and/or reverting unauthorized sharing of data within the network of business partners. Such perspective may be transversal to hierarches or managerial hubris (e.g. CIOs) or middle managers or specialized consultants (e.g. Big Data engineers/analysts or data forensic scientists).

References

Abrantes, B.F., Venkataraman, A. (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*. Retrieved from: https//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., & Lee, J. (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., Posch, P. and Somekh, B. (1996) *Teachers Investigate Their Work: An Introduction To The Methods Of Action Research*. London: Routledge

André, Q., Carmon, Z., Wertenbroch, K., Crum, A., Frank, D., Goldstein, W., Huber, J., Van Boven, L., Weber, B. and Yang, H. (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 April 5th, 2020)

Baruh, L. and Popescu, M. (2017) Big data analytics and the limits of privacy self-management, New media & society, Vol 19 No 4, pp. 579-596

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

Beulke, D. (2011) Big data impacts data management: The 5 Vs of big data, November 1st, 2011, available from https://davebeulke.com/big-data-impacts-data-management-the-five-vs-of-big-data, accessed December 6th 2019

Big Project (Online) BIG - *Big Data Public Private Forum – Objectives*. Retrieved from: https://www.big-project.eu/index.html (accessed March 19th 2021)

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

Braun, V., & Clarke, V. (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., Curry, E., & Wahlster, W. (2016). *New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe*. Springer Nature.

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

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

Crawford, K. and Schultz, J. (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., Papakonstantinou, V., Malgeiri, G., Beslay, L. and Sanchez, I. (2017) 'The right to data portability in the GDPR: towards user-centric interoperability of digital services', Computer Law & Security Review [online] https://ac.els-cdn.com/S0267364917303333/1-s2.0-S0267364917303333-main.pdf?_tid=c61e5bb2-11b8-464d-98a2-f01508a953cf&acdnat=1520597308_0f87eb7bf694fb7eeb6d95f7f20054df.

De Mauro, A., Greco, M. and Grimaldi, M. (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, available from https://www.demant.com/investor-relations/annual-report-2019, accessed May 7th 2020

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

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

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

EU (2020) EU data protection rules, available from 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 March 16th, 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 [online] 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. Retrieved from: https://edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en (Accessed on March 25th 2021)

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

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

Gartner (2019) Gartner Glossary, nd, available from https://www.gartner.com/en/information-technology/glossary/big-data (accessed December 5th 2019).

Hannak, A., Soeller, G., Lazer, D., Mislove, A. and Wilson, C. (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 López, P. (2011) The world's technological capacity to store, communicate, and compute information. *Science*, 332(6025), 60-65.

Hoagland, A.S. (2003) History of magnetic disk storage based on perpendicular magnetic recording, *IEEE transactions on magnetics*, 39(4), 1871-1875.

Gill, J. and Johnson, P. (2002) Research methods for managers, London, Sage.

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. Retrieved from; https://blogs.idc.com/2018/10/11/5-things-you-didnt-know-about-tech-spending/, accessed December 8th, 2019

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

Jacobs, A. (2009) The pathologies of big data. *Communications of the ACM*, 52(8), 36–44. doi:10. 1145/1536616.1536632

Kim, G.H., Trimi, S. and Chung, J.H. (2014) Big-data applications in the government sector, Communications of the ACM, Vol 57, No 3, p. 8

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

Leinweber, D.J. (2007) Stupid data miner tricks: overfitting the S&P 500. *The Journal of Investing*. 16(1), 15-22. Retrieved from: https://www.researchgate.net/profile/David_Leinweber2/publication/247907373_Stupid_Data_Miner_Tricks_Overfitting_the_SP_500/links/563bc8ca08ae405111a77817.pdf. (Accessed March 19th 2021).

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

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

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

Marz, N. and Warren, J. (2012) *Big Data: Principles and Best Practices of Scalable Real-time Data Systems*. MEAP edition. Westhampton, NJ: Manning.

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

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

Muhammad, S. S., Dey, B. L., & Weerakkody, V. (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 Weinlich, P. (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, pp. 97-117, Retrieved from: https://econpapers.repec.org/article/cupbuetqu/v_3a30_3ay_3a2020_3ai_3a1_3ap_3a97-117_5f5.htm

TechTerms (2012), What units of measurement are used for data storage?, retrieved from: https://techterms.com/help/data_storage_units_of_measurement , accessed February 2nd 2020.

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

Zwitter, A. (2014) Big data ethics, *Big Data & Society*, 1-6.