SYSTEMS & MANAGEMENT
Electronic magazine

# PRIVACY IN BIG DATA: OVERVIEW AND RESEARCH AGENDA

**Celina Rebello Silva**
celina.rebello@coppead.ufrj.br
Federal University of Rio de Janeiro – UFRJ, Rio de Janeiro, Rio de Janeiro, Brazil

**Elaine Maria Tavares Rodrigues**
elaine.tavares@coppead.ufrj.br
Federal University of Rio de Janeiro – UFRJ, Rio de Janeiro, Rio de Janeiro, Brazil

## ABSTRACT

This article aimed to map the academic production in the area of privacy in Big Data, for the Administration domain, raising the state of the art on the subject and some research opportunities from the gaps in the academic literature. Privacy is a question still open and increasingly threatened by the ubiquitous effect of data-generating devices such as mobile phones, sensors and computers, and new analytical techniques such as Big Data. The extensive bibliometric survey of academic production on the subject was carried out in the main reference databases and analysis of content allied to text mining, using R language. The results pointed not only to areas of concentration and research topics but also new alternatives of research in the area.

**Keywords**: Privacy; Big data; Data Mining.

PROPPI / LATEC
DOI: 10.20985/1980-5160.2017.v12n4.1205

## 1. INTRODUCTION

New devices and analytical techniques, such as the Internet of Things and Big Data, have recently expanded the capacity of new information and communication technologies. Sensors, social media and Radio Frequency Identification (RFID) tags, for example, have increased the already existing data overhead in organizations. As a way of dealing with the high volume of data, in different formats, new analytical techniques were developed.

Big Data is related to data sets, whose size is beyond the capacity of typical database software tools to capture, store, manage and analyze (Minelli et al., 2013; Ohlhorst, 2013). The main attributes related to the Big Data concept are volume, velocity and variety (Simon, 2013).

The volume is related to the growing amount of data, which directly impact organizational processes and influences predictive and statistical methods. In a highly competitive market or in highly complex administrative contexts, finding new ways to interpret data and process it faster has proven to be an important capability. The variety is related to the ability to analyze a wide range of data types and sources, including structured, semi-structured and unstructured data (Ohlhorst, 2013) which, as Big Data, take the form of messages, images and other types of data in social networks, sensors, GPS of cell phones, among others. Finally, speed refers to the ability to analyze data more quickly, sometimes in real time (McAfee et Brynjolfsson, 2012).

To these attributes is added the value - aggregated result from the analyzes of the information, such as the quality of the information or the financial value extracted (Beath et al., 2012; Maçada et Canari, 2014); and veracity - relative to the purity and authenticity of the information (Ohlhorst, 2013).

Like any technology embedded in the organizational context, the implementation and use of Big Data goes through challenges, some of them explored in the academic literature. One of the haziest points is that Big Data raises issues of concern to ethics, such as which data can be used in an analysis (Tene et Polonetsky, 2013). In this sense, some questions emerge, among which: under which conditions can one be considered as part of a large dataset? What if some post in the 'public' domain is taken out of context and analyzed in a way the author never imagined? What does it mean for someone to be identified, or to be analyzed without knowing (Boyd et Crawford, 2012)?

It is necessary to discuss when and what data can be considered as part of the Big Data strategy, given that the difficulty of ensuring the security and privacy of data can make projects unfeasible (Boyd and Marwick, 2011). It is essential to keep a constant ethical questioning not only in terms of use, but also in terms of collecting, storing and controlling access to such data (Simon, 2013; Tene et Polonetsky, 2013).

Considering that privacy is among the main concerns in Big Data, this article aims to map the academic production in Administration on privacy in this domain, revealing the state of the art on the topic and identifying opportunities for future research in gaps established in the literature. Through a literature review and the use of bibliometric techniques, the article raises issues and indicators related to the content of the publications and main focuses of the investigations, such as questions about public and private data, individual and collective, and how these aspects are addressed by researchers of the Administration area.

Bibliometric research acts as an enabler of the mapping of academic production on the theme of privacy in Big Data, metric by counting articles by periodicals and authors. In a second stage, the content of the mapped articles is mined in its entirety and analyzed.

The relevance of this study is reinforced by the fact that the concept of privacy is diffuse and the research on the topic of privacy in administration has very different biases. Some articles, such as Boyd and Crawford's (2012), address questions about Big Data in the social sciences; others, such as Martin's (2015), are focused on ethical issues. Chen et al. (2012), in turn, show the evolution of Big Data in the technological and strategic context for organizations. These works begin to address the unfolding of the privacy context in Big Data, but none elucidates the stage of scientific production achieved for the area.

The article is structured in four sessions. The first is the presentation of the concepts and definitions of the research themes and the second deals with the process of surveying the articles, research bases, treatment of the answers and analysis tools for content mining. The third session describes the analysis and interpretation of the results, pointing out research gaps to be explored; and the fourth presents the conclusions.

### 1.1 Big Data

Much is said about Big Data. Its economic and political potential, in addition to other dimensions, in which the phenomenon is implicated, are recognized. However, according to Boyd et Crawford (2012), Martin (2015) and Zuboff (2015), there is no consensus for the definition of the term. In this sense, Gandomi et Haider (2015) made the effort to establish an evolutionary line of definitions and their contents. Thus, according to the literature, Big Data is a phenomenon defined by several lenses.

Diebold (2012) places Big Data as a key subject for all sciences and claims to be his first term reference in an academic paper. However, recognizing that the origin of the term refers to non-academic work by computer science experts, Gandomi et Haider (2015) reaffirm the multifaceted character of the definition of Big Data, starting with Doug Laney's 3Vs model, which, despite not being academic, is cited in Chen et al. (2012). Gandomi et Haider (2015) believe that term definitions have evolved rapidly by the adoption incentive coming from companies such as IBM, EMC, Teradata and SAP, among other giants in the computing industry.

In the 3V model, volume refers to the amount of data collected and stored and, due to the magnitude, processing capacity was not yet available to act on them. Variety refers to the diversity of data nature: sensors, video files, tweets, images, among all forms of digital production available. Speed refers to the undeniable flow of information thanks to ubiquitous computing that often allows for real-time analysis. Gandomi et Haider (2015) point out that Laney's 3V model has evolved to 6V, where the fourth V is defined by IBM as veracity; the fifth V is defined by Oracle as the value; and the sixth V is defined by the SAS as variability. Veracity refers to how reliable the data is; the value refers to the value that this volume of data adds to the organization; and, finally, variability refers to changes in the data structure and how users interpret the same data.

From Laney's more traditional 3V model, where Big Data is defined by speed, variety and volume, to this day, Big Data's definitions multiply and gain a strategic focus. In Chen et al. (2012), the 3V model is associated with analytical techniques and evaluated by the authors as a blue ocean of opportunities for business, research and various applications. Big data is defined as the set of technologies and tools (databases and data mining tools), as well as techniques (analytical methods), capable of large-scale employment for complex data, in a universe of applications, in order to improve the performance of organizations. Storage, management, data analysis and visualization processes are part of the Big Data framework.

For Boyd et Crawford (2012), however, the definition of Big Data is described as: "a cultural, technological and academic phenomenon that establishes itself between the dynamics of technology, analysis and mythology that provokes utopian and dystopian rhetoric." For the authors, Big Data is a socio-technical phenomenon, whose real benefits must be critically questioned and carefully examined. They point out that Big Data is seen as a tool of high potential for social ills, such as terrorist cell identification, cancer cure and so on and, at the same time, threatening for its potential to hurt issues, such as the individual's right to privacy. This same critical view of unfolding Big Data usage can be seen in Martin (2015).

### 1.2 Privacy

The word *privacidade* (in Portuguese) is referred to as an Anglicism of privacy, which is rooted in the Latin term *privare*. As well as Big Data, the word privacy does not have an objective and unique concept. There are several doctrinal positions regarding its meaning, and can be descriptive or normative. Often, privacy is defined in terms of information control.

Privacy definitions are used to denote how people define situations and conditions of secrecy and how they are assessed, or to indicate the need for restrictions on the use of information or its processing. Newell (1995) analyzes privacy in a multidisciplinary perspective, defining forms and reasons for its establishment, understanding the term as temporary separation of the public domain. Glenn (1966) argues that there is no consensus as to whether or not privacy is a right between scholars of law and philosophy, and that at the time of its study, it was already controversial.

Informational privacy, in a normative sense, usually refers to a non-absolute moral right of persons to have direct or indirect control over access to (1) information about oneself, (2) situations in which others could acquire information about themselves, and (3) technology that can be used to generate, process, or disclose information about oneself.

The need to protect private life arose from the conflicting relationship between the individual and the order imposed by society. Nissembaum (1997) puts the definition of privacy by Charles Fried as: "Privacy is not simply the absence of information about us in the minds of others; on the contrary, it is the control we have over information about ourselves." The author follows the definition of W.A. Parent: "Privacy is the condition of a person not having their personal information irregularly known by others," referring to facts that most people in a given society choose not to disclose about themselves (except for friends, family, advisors, etc.) or to facts about which a particular person is extremely sensitive, choosing not to reveal. Westin (2003), on the other hand, says that privacy is the assertion that individuals and groups determine for themselves when, how, and to what extent information about them is communicated to others. Beardsley (1971) suggests that people have the right to decide when and how much information about them will be revealed to others. Gavison (1980), however, has offered variants of privacy definitions in three spheres, all as a central aspect of access restriction: the first as a right, the second as a loss of privacy and the third with punishments. The central idea is the issue of limiting data and information to people, or information about people.

Warren et Brandeis (1890) already protested against intrusive journalistic activities. At the time, they raised the debate about the individual limits in terms of how far society could know something about a citizen. They emphasized that political, social and economic changes demand the recognition of new rights, and the common law continues to evolve in order to guarantee the new demands of society. The law, according to the authors, is updated in order to restore balance and inviolability of fundamental rights.

Laws try, through doctrines, to limit the invasiveness of some privacy displays. We have in Fuster (2014) the question of privacy in the light of Robert Alexy's sphere theory. This theory reinforces the concepts of Parent (1983), defining the existence of three different levels of protection of private life, spheres with different protection intensities. It assumes the **innermost sphere** (ultimate intangible scope of human freedom) as the most intimate, intangible, extremely reserved, with matters that are most secret and that should not come to the knowledge of others; the **private broad sphere** as the private scope insofar as it does not belong to the innermost sphere, including subjects that individuals bring to the knowledge of another person of their trust, excluding the rest of the community; and the **social sphere**, which encompasses everything not included in the broad private sphere: all matters related to the news that the person wishes to exclude from the knowledge of others. The right to privacy, then, would be defined as what preserves us from the knowledge of others, reserving our own experience.

There are moral reasons for protecting personal data and for providing direct or indirect control over access to this data by third parties. Van den Hoven (2008) defines them as:

- Damage prevention: unrestricted access by others to an individual's passwords, where features and whereabouts can be used to harm the data holder in a variety of ways;

- Asymmetry of information: personal data have become commodities. Typically, individuals are not in a good position to negotiate contracts on the use of their data and do not have the means to verify whether the partners will follow the terms of the contract. Data protection laws, regulations and governance are designed to establish fair conditions for the drafting of contracts on the transmission of personal data, exchange and provision of data with brakes, counterweights and repair guarantees;

- Informational injustice and discrimination: Personal information provided in one sphere or context (e.g. health care) can change its meaning when used in another sphere or context (such as business transactions) and can lead to discrimination and harm to individuals;

- Invasion of moral autonomy: lack of privacy can expose individuals to external forces that influence their choices.

To the right, all these formulations provide good moral grounds for limiting and restricting access to personal data, giving individuals control over their data. When third parties take an invasive stance, episodes such as the Carolina Dieckman case, occurred in Brazil, which resulted in a law; and, worldwide, the revelations of Edward Snowden and their implications in economic and diplomatic relations.

In order to find this more intimate and internal scope of the individual, the existence of some set of behaviors that respects the interests of community life is questioned. Perhaps the nature of the social structure has developed in such a way that the recent past forces the recognition that privacy, hitherto presumed as an ingredient of moral action, should now be specified as a right. The philosophy that describes the political structure as essentially corporate in nature has traditionally derived, or resulted in a description of the moral status of the individual that not only denies the right to privacy, but designates a political and moral offense.

Warren et Brandeis (1890) argue that the act of publishing certain content causes individuals to give up their right to privacy. On the other hand, as slander is generated by third parties, it must be treated in accordance with the legal instruments. Despite Warren et Brandeis' (1890) argumentative efforts, Glenn (1966) discusses the contrast between definitions of privacy from Hegel's *moralität*[1] distinction, as relative to private judgment, and *Sittlichkeit*[2], as obligations defined by corporate and institutional orders, stating that the claim of privacy is simply triviality, because of irresponsibility practiced by individuals. He continues his analysis on privacy by pointing to individuals as entirely dependent on the degree to which they identify their interests and rights with the appropriate value structure for the private and corporate order in which they find themselves. Since the rights and duties of individuals are determined by the existing orders, in which they also participate and whose highest form is the state, then the reduction of privacy or its limit is realized, with the claim that, in the last instance, the individual must accept the interpretation of a "judge" who correctly discerns values, duties and obligations of the historical

[1] T.N.: *Moralität* – Morality, in English.

[2] T.N.: *Sittlichkeit* – Morality, in English.

moment.

Glenn (1966) establishes the counterpoint to Anglo-American philosophy, since it defines the political structure as collective legitimacy on which privacy depends. It derives and depends on individual judgments of those who are constituted in that community, a definition that is aligned with the definition of Newell (1995). He perceives the problems that could arise from confusion between moral responsibility and legal responsibility. Privacy is assumed to be a right justified by utility, if not by nature, an essential ingredient of Anglo-American political philosophy, like any right, that must be protected by law. The persistence of the ill-defined assumption for the idea of privacy opens dangerous and threatening precedents and forms a gray zone between moral responsibility and legal responsibility arising from the dilemma that invasion of privacy should be punished and even for reasons purely utilitarian, whether punishment would do more harm than good.

The question of privacy, in Glenn (1966), is contextualized considering geographic, moral and religious heterogeneity, in counterpoint to the contextualization of Hegel. Hegel sought the ambiguity of historical dialectics to describe a political organism of such homogeneity, whose societal structure would resolve relations, that is, moral, political, individual, and social conflicts. The Hegelian model falls into disuse because of the complexity of individual issues, through private judgment and opposition. The face-to-face society, a Greek heritage, rejected the concept of privacy.

Glenn (1966) points out that Bentham recognized that the moral status of the individual needed protection against an increasingly intrusive and dominant social organization. The rejection of privacy took the form of aggressive attacks, culminating in private judgments and opinion. Hegel succeeded in describing the nature of the political order that was rapidly becoming real and, even though its premises were unacceptable, much progress was made by his analysis of the institutional structure of the social order. Bentham, in refusing to abandon his system of individual values, demonstrated that traditional patterns of political analysis were inadequate to formulate the status of the individual in the political and legal contexts of society and the state of the time.

### 1.3 Privacy on Big Data

The dynamics of privacy in Big Data have similar points about the challenges of Warren et Brandeis (1890) about the ease with which information is disclosed. However, Matzner (2014) warns that, currently, people make available data without any criteria, accepting as a counterpart, benefits of little value or nothing in return.

Kshetri (2014) and Zuboff (2015) point out that the ubiquity of Big Data also favors the exacerbation of power asymmetries between states, industries, groups and individuals. The trends are the growth of privacy problems, due to the production of data and the exposure of information of private content, collected without full awareness of individuals. McNeely et Hahm (2014) ask questions such as: what data are collected and which are not? Why? What is used and what is not? What are the implications of this selection? How and why? What is not measured? What fundamental or critical factors should be considered for full understanding of a particular phenomenon or condition? These issues point to the need for a critical approach to Big Data in terms of understanding its essence, usage and effects.

This same class of questions is approached by Martin (2015), when analyzing the supply chain characteristic of Big Data, of the positive and negative uses of the technology, emphasizing aspects such as resale of data and the risk of misuse of information, with significant impact on users, such as destruction of value, reduction of rights of interested parties, and disrespect to any individual involved in the process. The author further argues that the Big Data industry generates negative aggregate externality by expanding the surveillance system through which a range of information is collected and gathered invisibly to the user.

The issues of Boyd et Crawford (2012), coupled with the technical and environmental aspects pointed out by Chen et al. (2012) and by Martin's (2015) strategic vision, one realizes that the reality provided by the Big Data phenomenon generates implications on the privacy issue, which cannot be ignored or neglected. The technical aspect appears as an environmental condition, requiring the contextualization of its use and applications, and a profound analytical approach is needed, not only from the perspective of the Big Data chain, but also from a step beyond: the establishment of rules and outlines, definitions of limits so that this reality, called Big Data, generates positive externalities and not what has already been perceived recently as the practices denounced by Edward Snowden. At the current stage of the dynamics, it is necessary to establish requirements that balance the relationship between access to information and the right to privacy.

The analytical requirements demanded by privacy problems are basically two: the definition of values and the specification of procedures. The first is the moral task of philosophy, reflected in legislation; the second depends on the legal process, especially if it operates in the control of administrative functions in which the individual is subject. Failure or negation in performing these tasks leaves us with the only alternative of increasing arbitrary administrative control that can actively achieve individual values. The control of legal proceedings is mandatory, but must be justified. If privacy is defined as an essential requirement for the attainment of morality, then privacy is a right that the law should not only protect, but provide. Modern man is born chained and only the law could free him. What is perceived, and will be shown later, is a strong trend of research on public policy and digital politics that addresses the still undetermined limits of the dynamics of privacy in Big Data.

## 2. BIBLIOMETRIC RESEARCH

The beginning of the bibliometric process was the choice of the bases and the definition of the reference control tools. For reference control, the Zotero software was used because of its ability to integrate with web browsers, as well as for the functionalities of redundancy treatment and interoperability with "RIS" formats, helping to classify categories and eliminate redundancies.

The search was essentially for the combination of the key words "Big Data" and "Privacy" in the title, summary or keywords field. The criterion was: the search should only be carried out in academic journals reviewed by specialists, with full text availability in the respective databases, from 2000 to 2015.

The databases were selected from September to October 2015, in order to obtain coverage that is believed to be reasonable for a robust literature review. Articles without authors and anonymous have been withdrawn. Preliminary results without detailed reading of abstracts presented production published in the following databases: Ebsco, Emerald Insight, JStor, Proquest, Sage, Science Direct, Scopus, SpringerLink, Web of Science, and Wiley Online.

Some considerations were made for pre-selection of articles. Not all databases have the same functionality in their respective search interfaces; therefore, manual refinement of results is required. In the case of Springer, the discipline filters applied were: "Business and Management" and "Social Sciences". Nevertheless, articles such as interviews, research notes, and others that do not have a scientific article structure were returned. As for Wiley Online, it was not possible to select the basis for the articles of administration, and refinement was manually made on all results returned by session title of academic journals of Administration.

### 2.1 Results by base (2000-2015)

Publications dealing with privacy in Big Data show a peak in academic production in 2014, within the set of publications in Administration, as shown in figure 1.

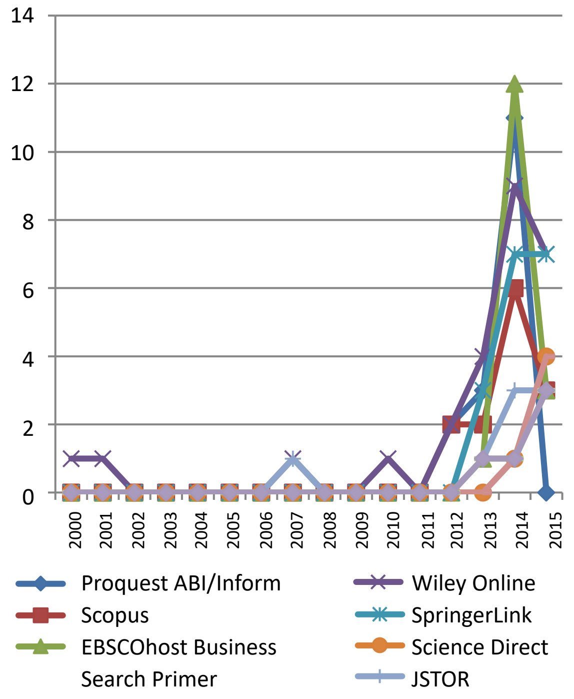

**Figure 1.** Time series of the number of publications in the consulted databases

**Source:** Authors (2015).

The elimination criteria were the collection of anonymous articles, without references in academic journals, redundant articles, with presence in more than one base, as well as reading the articles themselves. Articles that presented some of the mentioned characteristics were discarded for the phase of data mining. Articles whose references did not contain revised academic journals were also eliminated. Duplicate files have been deleted because of the prerequisites of the statistical data mining process, because redundancy generates mathematical distortions, thus leading to distorted results of the research, themes and areas of concentration. The result of the process of selecting the articles to be mined verbatim is presented in table 1, below.

As data mining is the bibliometric support employed, the protocol describes the process of word association, and with this, redundancy would cause greater weights to terms over others.

Among the bases researched, the greatest occurrence of redundancy occurred among those that have a portfolio with greater diversification of journals, that is, bases that

**Electronic Journal of Management & System**
Volume 12, Number 4, 2017, pp. 491-505
DOI: 10.20985/1980-5160.2017.v12n4.1205

**Table 1.** Result of the process of selecting the articles to be mined verbatim

| | Number of Articles | | | |
|---|---|---|---|---|
| **Bases** | **Iniciais** | **Redundantes** | **No references *** | **Final result** |
| EBSCO | 16 | 8 | 4 | 4 |
| EmeraldInsight | 5 | 0 | 0 | 5 |
| JSTOR | 8 | 0 | 0 | 8 |
| Proquest | 16 | 2 | 7 | 7 |
| Sage | 5 | 0 | 0 | 5 |
| Science Direct | 5 | 0 | 0 | 5 |
| Scopus | 7 | 3 | 1 | 3 |
| SpringerLink | 17 | 0 | 1 | 16 |
| Web of Science | 5 | 0 | 0 | 5 |
| Wiley Online | 22 | 0 | 11 | 11 |
| Total | 116 | 13 | 24 | 69 |

**Source:** Authors (2015)

aggregate the production of publishers. The intersections occurred essentially with EBSCO, Proquest and Scopus. The bases that are publishers, rather than the aggregating bases, were taken as a criterion of permanence..

## 3. DATA MINING AND CONTENT ANALYSIS

For the data mining process, a source code was developed in R language. At the end of the article selection process, 69 articles of peer reviewed publications were selected. Mining consisted of three steps: the conversion of files to mining; content analysis and statistical results; and visual analysis of statistical results. Data mining aims to identify correlations between key terms of privacy studies in Big Data in the domain of administration, and was performed considering English dictionary in successive stages, in order to eliminate terms of high frequency of occurrence with low semantic significance, namely: connectives, prepositions, pronouns, interjections and adverbs. In this way, it was possible to observe in the second phase of mining that only the classes of words that are nouns, adjectives, and verbs, with low semantic load and relevance were removed, using the dictionary of semantic associations of terms. In the third phase, only low frequency words were computed, but with high semantic load and relevance.

After the mined body was essentially formed of nouns, adjectives and verbs, we performed the frequency analysis of terms and the correlation between terms with semantic association with the keywords, component words of the domain privacy and Big Data, and with those of greater semantic significance in the texts returned by the bases.

After mapping these relationships, the order in which the terms occurred in quantity and relevance was calculated, as well as their association in clusters in a second step. The clusters association assisted in the visual identification of the correlations between the components of the studies, as well as in the identification of clusters

### 3.1 Analysis of terms and frequencies

The third phase of data mining, presented in figure 2, found the following results: *"social", "security", "access", "approach", "context", "control", "government", "development", "disclosure", "personal", "individuals", "google", "business", "facebook", "model", "mobile", "need", "personalization", "process", "services", "terms", "value", "risk", "scale", "role", "trust", "questions", "regulation".*

This set indicates that privacy-related research objects in Big Data go through issues pertaining to approaches, deployments, values, and rules, rather than technology itself. The order of magnitude with which these terms can be seen, according to figure 2, reveals that social aspects are the most important within the production of privacy in Big Data in the area of administration. Academic production has turned its attention to the roles of actors such as the individual, government and companies; control and access policies; processes, services and terms proposed to individuals by companies such as Google and Facebook; processes of treatment, custody, exposure, aperture, data extraction, risk and the future related to this context.

The complementation of the analysis by dendrogram representation (Figure 3) reinforced the suggestion that words such as "social", "personal" and "security" should be eliminated to deepen the next iteration. The elimination of these terms should favor the better understanding between word relations by the proximity of their levels in the structure. Once the terms have been eliminated, new arrangements of words should appear, with low frequency and strong semantic meaning, meaning the stage of research in the area of concentration, as well as their combination in new arrangements, potential indicators of research gaps.

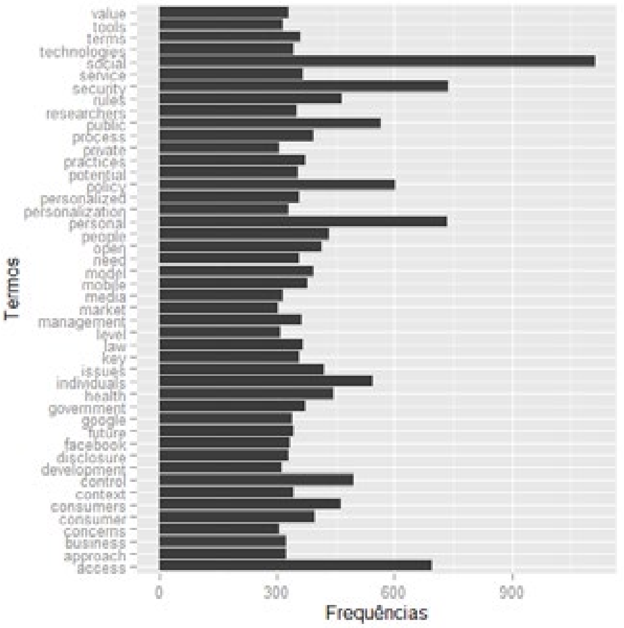

**Figure 2.** Analysis of terms and frequencies - first mining of the third stage

**Source:** Authors (2015)

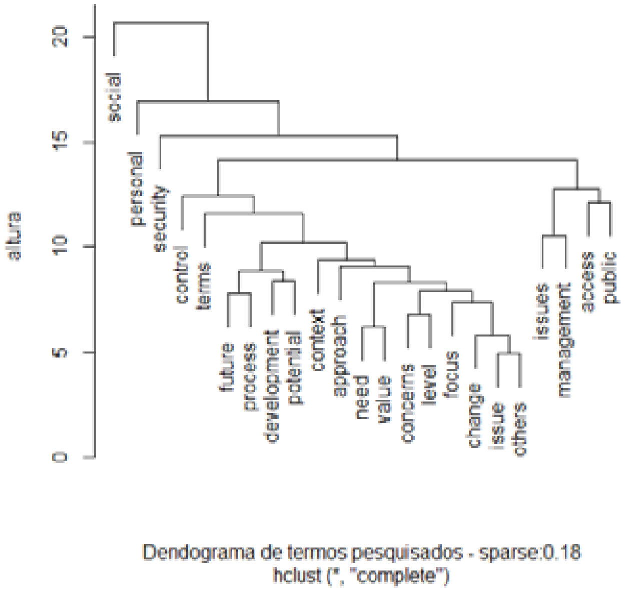

**Figure 3.** Dendrogram of the first mining of the third stage

**Source:** Authors (2015)

Thus, terms of low frequency and high semantics will be inputs and tend to participate in the next dendrogram. They should form new clusters, after the removal of the terms "social," "personal," and "security." This analysis generated Table 2, below:

Eliminating the terms suggested by the previous analysis, one has the second level of mining results. In it, it is possible to infer some terms with very close frequencies, such as "terms", "surveillance", "rules", "services", in a frequency group close to 400 occurrences, while two other groups contain "default", "control" , and a third one with "individuals", "management", "issues", "health", in the bar chart of figure 4, and in the dendrogram of figure 5.

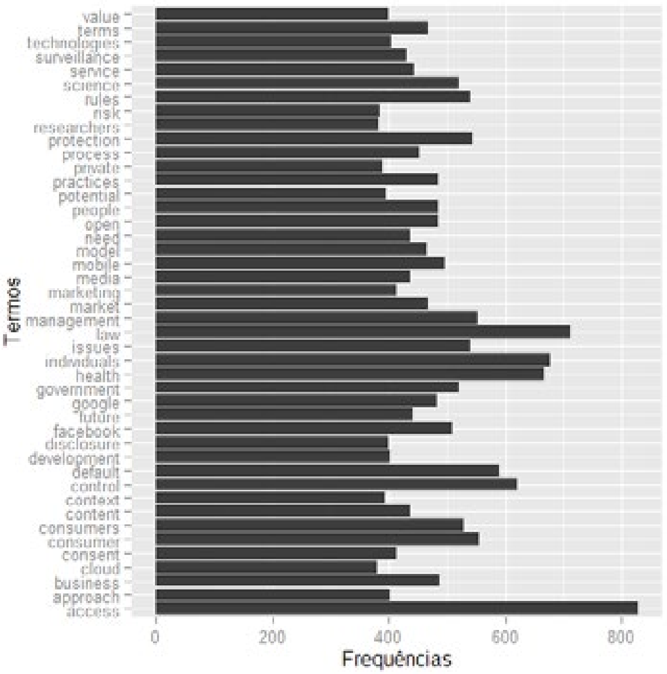

**Figure 4.** Analysis of terms and frequencies - first mining of the third stage.

**Source:** Authors (2015)

In this way, the result of the first mining generated a new dendrogram whose structure reflects not only the terms of table 2, but also points the indications to areas of concentration, such as for research on terms of control over people and access issues . The other branches of the dendrogram indicated research strands that consider the balance between the scientific approach of the privacy phenomenon in Big Data with the analysis of needs and value; knowledge of process contexts; and the development of future potential, in contrast to managerial aspects. Further interpretations and arrangements within the branches, such as the grouping suggested below, are possible in the dendrogram of Figure 6.

**Electronic Journal of Management & System**
Volume 12, Number 4, 2017, pp. 491-505
DOI: 10.20985/1980-5160.2017.v12n4.1205

**Table 2.** Analysis of low frequency terms (n=280).

| access | analytics | tools | business | challenges | cloud | collection | companies |
|---|---|---|---|---|---|---|---|
| people | concerns | model | policies | consumers | consent | content | researchers |
| control | default | terms | economic | Facebook | future | network | disclosure |
| mobile | industry | health | market | Context | key | media | knowledge |
| Law | process | trust | marketing | protection | issues | personalized | personalization |
| Need | condition | news | open | Google | sharing | government | management |
| press | potential | power | practices | Political | private | individual | development |
| quality | records | value | risk | Rules | science | service | technologies |
| Uses | society | Subject | support | approach | results | computing | surveillance |

Source: Authors (2015)

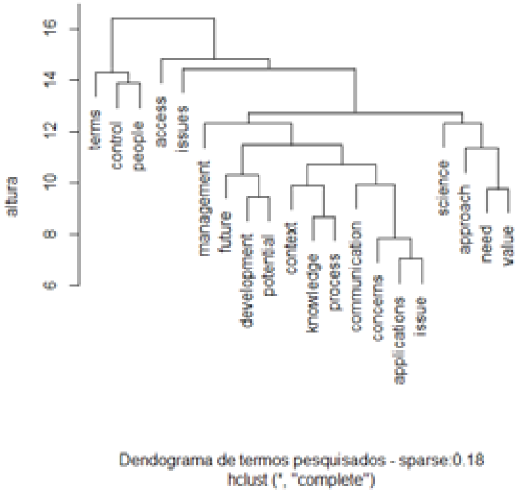

**Figure 5.** Dendrogram of the second mining of the third stage.
**Source:** Authors (2015)

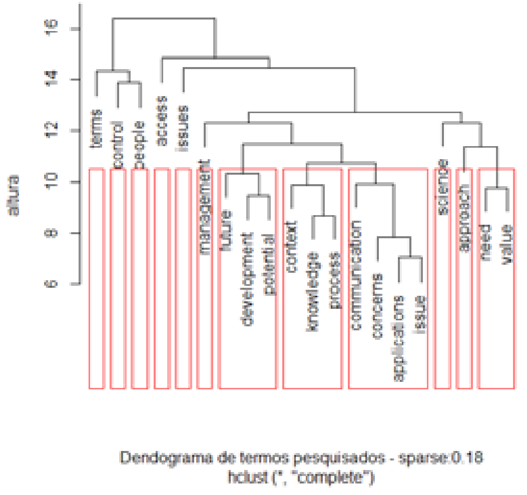

**Figure 6.** Groupings for potential research
**Source:** Authors (2015)

From the Dendrogram of Figure 6, it was possible to visualize some gaps for research, as counterpoints between the scientific approach and management approach to privacy issues in Big Data; questions of communication and contextualization between uses of the applications and processes in terms of who would be in possession of the knowledge; how people behave when controlled; and what is considered a necessity in terms of the context of Big Data and privacy, among other options.

The word map analysis in Figure 7, constructed from data mining, reveals the main issues addressed by the academic production of privacy focused on Big Data, ranging from economic relations on information exposure, to effects on governments , individuals and the relations of power and trust between these actors, and even rules and rights in the context of using social networks. It is interesting to note that its result expresses what was statistically raised in the previous steps of this study: the central issue of privacy in Big Data is the balance between access to the information of individuals, what rights are and how laws are treated, deal with these points.

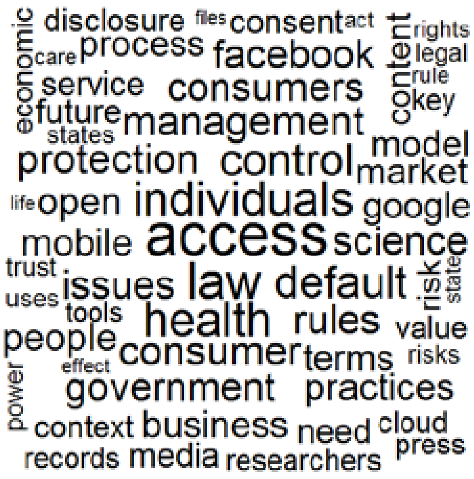

**Figure 7.** Map of words. Relevance is expressed by the visual magnitude of the terms
**Source:** Authors (2015)

## 4. FINAL CONSIDERATIONS

Digital devices transmit diverse information: some transparent to users, others loaded and shared by them as actors in the system. However, at some point in the data acquisition process, there is always human intervention to define what is transmitted or what use will be given for that information; and whether accepting contractual terms and conditions or leaving a device exposed or vulnerable, the effects of that practice befall on the individual. They are decisions taken at the corporate or governmental level or by individuals, which, indiscriminately, without limits, can have threatening implications, as shown by Martin (2015) and Zuboff (2015), thus eroding economic systems. Literature and analysis point out that this aspect makes up the core of Big Data privacy research challenges, along with the role played by organizations in the industry.

Big data is a reality and privacy issues have always existed. The field is not only a source of opportunities, but at the same time it is challenging for new research proposals, precisely in the sense of deepening the debate on ideas of socio-technical policies that encourage the positive use of Big Data, preserving privacy and generating positive externalities to individuals and other actors in this environment. The academic production analyzed focuses on questions of ethics and processes, with a greater focus on questioning, to the detriment of pointing out solutions.

## REFERENCES

Beardsley, E.L. (1971), “Privacy: Autonomy and Selective Disclosure”, in Pennock, J.R.; Chapman, J.W. (ed.), Privacy: Nomos XIII, Atherton Press, New York, pp. 56-70.

Beath, C.; Becerra-Fernandez, I.; Ross, J, et al. (2012), “Finding value in the information explosion”, MIT Sloan Management Review, Vol. 53, No. 4, pp. 18-20.

Boyd, D.; Crawford, K. (2012), “Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon”, Information Communication and Society, Vol. 15, No. 5, pp. 662-679.

Boyd, D.; Marwick, A. (2011), “Social privacy in networked publics: teens attitudes, practices, and strategies”, Privacy Law Schoolars Conference, 2 jun. 2011, disponivel em: http://www.danah.org/papers/2011/SocialPrivacy-PLSC-Draft.pdf (acesso em 20 jul. 2014).

Chen, H.; Chiang, R.H.L.; Storey, V.C. (2012), “Business Intelligence and Analytics: from big data to big impact”, MIS Quarterly Executive, Vol. 36, No. 4, pp. 1165-1188.

Diebold, F.X. (2012), “A Personal Perspective on the Origin(s)and Development of ‘Big Data’: The Phenomenon, the Term, and the Discipline, Second Version”, PIER Working Paper No. 13-003, disponível em: http://ssrn.com/abstract=2202843 (acesso em 11 nov. 2017).

Fuster G.G. (2014), The Emergence of Personal Data Protection as a Fundamental Right of the EU, Springer International Publishing, Switzerland.

Gandomi, A.; Haider, M. (2015), “Beyond the hype: Big data concepts, methods, and analytics”, International Journal of Information Management, Vol. 35, No. 2, pp. 137-144.

Gavison, R. (1980), “Privacy and the limits of the law”, The Yale Law Review Journal, Vol. 89, No. 3, pp. 421-471.

Glenn, N. (1966), “Philosophical Views on the Value of Privacy”, Law and Contemporary Problems, Vol. 31, pp. 319-325.

Kshetri, N. (2014), “Big data’s impact on privacy, security and consumer welfare”, Telecommunications Policy, Vol. 38, No. 11, pp.1134–1145.

Maçada, A.C.G.; Canary, V.P. (2014), “A Tomada de Decisão no Contexto do Big data: estudo de caso único”, XXXVIII Enanpad 2014, Rio de Janeiro, 13-17 set. 2014.

Martin, K.E. (2015), “Ethical Issues in the Big Data Industry”, MIS Quarterly Executive, Vol. 14, No. 2, pp.67–85.

Matzner, T. (2014), “Why privacy is not enough privacy in the context of “ubiquitous computing” and “big data””, Journal of Information, Communication Ethics in Society, Vol. 12, No. 2, pp. 93-106.

Mcafee, A.; Brynjolfsson, E. (2012), “Big Data: The Management Revolution”, Harvard Business Review, Vol. 90, No. 10, pp. 60-68.

Mcneely, C.L.; Hahm, J. (2014), “The Big (Data) Bang: Policy, Prospects, and Challenges”, Review of Policy Research, Vol. 31, No. 4, pp. 304–310.

Minelli, M.; Chambers, M.; Dhiraj, A. (2013), Big Data Big Analytics: Emerging Business Intelligence and Analytic trends for today’s businesses. John Wiley & Sons, Hoboken, New Jersey.

Newell, P. (1995), “Perspectives on Privacy”, Journal of Environmental Psychology, Vol. 15, No. 2, pp. 87–104.

Nissembaum, H. (1997), “Toward an approach to privacy in Public: Challenges of Information Technology”, Ethics and Behavior, Vol. 7, No. 3, pp. 207-219.

Ohlhorst, F. (2013), Big Data Analytics: Turning Big data into Big money, [S.l.]: Wiley.

Parent, W.A. (1983), ”Privacy, Morality and the Law”, Philosophy and Public Affairs, Vol. 12, No. 4, pp. 269-288.

Simon, P. (2013), Too Big too Ignore, John Wiley & Sons, Hoboken, New Jersey.

Tene, O.; Polonetsky, J. (2013), "Big Data for All: Privacy and User Control in the age of Analytics", Northwestern Journal of Technology and Intellectual Property, Vol. 11, No. 5.

Van den Hoven, J.; Weckert, J. (2008), "Information technology, privacy, and the protection of personal data", Information Technology and Moral Philosophy, Cambridge University Press, Cambridge.

Warren, S.D.; Brandeis, L.D. (1890), "The Right to Privacy", The Harvard Law Review, Vol. 4, No. 5, pp. 193-220.

Westin, A.F. (2003), "Social and Political Dimensions of Privacy", Journal of Social Issues, Vol. 59, No. 2, pp.431-453.

Zuboff, S. (2015), "Big other: Surveillance capitalism and the prospects of an information civilization", Journal of Information Technology, Vol. 30, No. 1, pp. 75–89.

**Received:** July 22, 2016.

**Approved:** Oct 11, 2017.

**DOI:** 10.20985/1980-5160.2017.v12n4.1205

**How to cite:** Silva, C.R.; Rodrigues, E.M.T. (2017), "Privacy in Big Data: overview and research agenda", Sistemas & Gestão, Vol. 12, No. 4, pp. 491-505, available from: http://www.revistasg.uff.br/index.php/sg/article/view/1205 (access day abbreviated month. year).

**Annex 1 - Articles raised**

The following articles were the basis for data mining:

Ambrose, M.L.; Ausloos, J. (2013), “The Right to Be Forgotten Across the Pond”, Journal of Information Policy, Vol. 3, pp.1-23.

Barbu, A. (2013), “Eight contemporary trends in the market research industry”, Management& Marketing, Vol. 8, No. 3, pp. 429-450.

Barocas, S.; Nissenbaum, H. (2014), “Big Data’s End Run Around Procedural Privacy Protections”, Communications of the ACM, Vol. 57, No. 11, pp. 31–33.

Bates, D.W.; Saria, S.; Ohno-Machado, L., et al. (2014), “Big data in health care: Using analytics to identify and manage high-risk and high-cost patients”, Health Affairs, Vol. 33, No. 7, pp. 1123-1131.

Bertot, J.C.; Gorham, U.; Jaeger, P.T., et al. (2014), “Big data, open government and e-government: Issues, policies and recommendations”, Information Polity, Vol. 19, No. 1/2, pp. 5–16.

Bozdag, E. (2013), “Bias in algorithmic filtering and personalization”, Ethics and Information Technology, Vol. 15, No. 3, pp. 209–227.

Bragge, J.; Sunikka, A.; Kallio, H. (2012), “An exploratory study on customer responses to personalized banner messages in the online banking context”, Journal of Information Technology Theory and Application, Vol. 13, No. 3, pp. 5-18.

Brennan, N.; Oelschlaeger, A.; Cox, C., et al. (2014), “Leveraging the big-data revolution: CMS is expanding capabilities to spur health system transformation”, Health Affairs, Vol. 33, No. 7, pp. 1195-1202.

Cate, F.H.; Mayer-Schönberger, V. (2013), “Notice and consent in a world of big data”, International Data Privacy Law, Vol. 3, No. 2, pp. 67-73.

Cate, F.H.; Cate, B.E., “The supreme court and information privacy”, International Data Privacy Law, Vol. 2, No. 4, pp. 255-267.

Chow-White, P.A.; Macaulay, M.; Charters, A., et al. (2015), “From the bench to the bedside in the big data age: ethics and practices of consent and privacy for clinical genomics and personalized medicine”, Ethics and Information Technology, Vol. 17, No. 3, pp. 189–200.

Chung, T.S.; Wedel, M.; Rust, R.T. (2015), “Adaptive personalization using social networks”, Journal of the Academy of Marketing Science, Vol. 44, No. 1, pp. 66-87.

Cohen, I.G.; Amarasingham, R.; Shah, A., et al. (2014), “The legal and ethical concerns that arise from using complex predictive analytics in health care”, Health Affairs, Vol. 33, No. 7, pp. 1139-1147.

Cranor, L.F.; Sadeh, N. (2013), “Privacy engineering emerges as a hot new career”, IEEE Potentials, Vol. 32, No. 6, pp. 7–9.

Crosas, M.; King, G.; Honaker, J., et al. (2015), “Automating Open Science for Big Data”, The ANNALS of the American Academy of Political and Social Science, Vol. 659, No. 1, pp. 260–273.

Curtis, L.H.; Brown, J.; Platt, R. (2014), “Four health data networks illustrate the potential for a shared national multipurpose big-data network”, Health Affairs, Vol. 33, No. 7, pp. 1178-1186.

Daries, J.P.; Reich, J.; Waldo, J., et al. (2014), “Privacy, Anonymity, and Big Data in the Social Sciences”, Communications of the ACM, Vol. 57, No. 9, pp.56–63.

Duan, R.; Hong, O.; Ma, G. (2014), “Semi-Supervised Learning in Inferring Mobile Device Locations”, Quality and Reliability Engineering International, Vol. 30, No. 6, pp. 857–866.

Einav, L.; Levin, J. (2014), “The Data Revolution and Economic Analysis”, Innovation Policy and the Economy, Vol. 14, No. 1, pp. 1–2.

Fabian, B.; Ermakova, T.; Junghanns, P. 92015), “Collaborative and secure sharing of healthcare data in multi-clouds”, Information Systems, Vol. 48, pp. 132–150.

Fleurence, R.L.; Beal, A.C.; Sheridan, S.E., et al. (20140, “Patient--powered research networks aim to improve patient care and health research”, Health Affairs, Vol. 33, No. 7, pp. 1212-1219.

Frizzo-Barker, J.; Chow-White, P. (2014), “Research in brief: From patients to petabytes: Genomic big data, privacy, and informational risk”, Canadian Journal of Communication, Vol. 39, No. 4, pp. 615-625.

Gehrke, J. (2012), “Quo vadis, data privacy?”, Annals of the New York Academy of Sciences, Vol. 1260, No. 1, pp. 45–54.

Genov, N. (2015), “The future of individualization in Europe: changing configurations in employment and governance”, European Journal of Futures Research, Vol. 2, No. 1, pp. 1–9.

Gleibs, I.H. (2014), “Turning Virtual Public Spaces into Laboratories: Thoughts on Conducting Online Field Studies Using Social Network Sites”, Analyses of Social Issues and Public Policy, Vol. 14, No. 1, pp. 352–370.

Habte, M.L.; Howell, C.; Warren, A. (2015), “The Big Data Dilemma: Compliance for the Health Professional in an Increasingly Data-Driven World”, Journal of Health Care Compliance, Vol. 17, No. 3, pp. 5–12.

Haggerty, K.D.; Ericson, R.V. (2000), “The surveillant assemblage”, The British Journal of Sociology, Vol. 51, No. 4, pp. 605–622.

Hardin, S. (2013), “ASIS&T annual meeting plenary speaker: Edward chang: Mobile opportunities”, Bulletin of the American Society for Information Science and Technology, Vol. 39, No. 3, pp. 46–48.

Heffetz, O.; Ligett, K. (2014), "Privacy and Data-Based Research", The Journal of Economic Perspectives, Vol. 28, No. 2, pp. 75–98.

Helles, R.; Lomborg, S. (2013), "Regulatory response? Tracking the influence of technological developments on privacy regulation in Denmark from 2000 to 2011", Policy & Internet, Vol. 5, No. 3, pp. 289–303.

Hirsch, P.B. (2013), "Corporate reputation in the age of data nudity", Journal of Business Strategy, Vol. 34, No. 6, pp. 36–39.

Hofman, W.; Rajagopal, M. (2014), "A technical framework for data sharing", Journal of Theoretical and Applied Electronic Commerce Research, Vol. 9, No. 3, pp. 45–58.

Hogan, M.; Shepherd, T. (2015), "Information Ownership and Materiality in an Age of Big Data Surveillance", Journal of Information Policy, Vol. 5, pp. 6–31.

Holt, J.; Malčić, S. (2015), "The Privacy Ecosystem: Regulating Digital Identity in the United States and European Union", Journal of Information Policy, Vol. 5, pp. 155–178.

Hull, G. (2015), "Successful failure: what Foucault can teach us about privacy self-management in a world of Facebook and big data", Ethics and Information Technology, Vol. 17, No. 2, pp. 89–101.

Johnson, J.A. (2014), "From open data to information justice", Ethics and Information Technology, Vol. 16, No. 4, pp. 263–274.

Krishnamurthy, R.; Desouza, K.C. (2014), "Big data analytics: The case of the social security administration", Information Polity: Vol. 19, No. 3/4, pp. 165–178.

Kshetri, N. (2014), "Big data's impact on privacy, security and consumer welfare", Telecommunications Policy, Vol. 38, No. 11, pp. 1134–1145.

Kwon, O.; Lee, N.; Shin, B. (2014), "Data quality management, data usage experience and acquisition intention of big data analytics", International Journal of Information Management, Vol. 34, No. 3, pp. 387-394.

Kyunghee Y.; Hoogduin, L.; Zhang. (2015), "Big Data as Complementary Audit Evidence", Accounting Horizons, Vol. 29, No. 2, pp. 431–438.

Laat, P.B. (2014), "From open-source software to Wikipedia: "Backgrounding" trust by collective monitoring and reputation tracking", Ethics and Information Technology, Vol. 16, No. 2, pp.157–169.

Leonard, P. (2014), "Customer data analytics: Privacy settings for 'big data' business", International Data Privacy Law, Vol. 4, No. 1, pp. 53-68.

Lesley, W.S.; Shmerling, S. (2015), "Risks and Opportunities of Data Mining the Electronic Medical Record", Physician Leadership Journal, Vol. 2, No. 4, pp. 40–45.

Libaque-Sáenz, C.F.; Wong, S.F.; Chang, Y., et al. (2014), "Understanding antecedents to perceived information risks an empirical study of the Korean telecommunications market", Information Development, Vol. 32, No. 1, pp. 1-16.

Liu, D.; Wang, S. (2013), "Nonlinear order preserving index for encrypted database query in service cloud environments", Concurrency and Computation: Practice and Experience, Vol. 25, No. 13, pp. 1967–1984.

Liu, Y. (2014), "User control of personal information concerning mobile-app: Notice and consent?", Computer Law & Security Review, Vol. 30, No. 5, pp. 521–529.

Martin, K.E. (2015), "Ethical Issues in the Big Data Industry", Mis Quarterly Executive, Vol. 14, No. 2, pp. 67–85.

Matzner, T. (2014), "Why privacy is not enough privacy in the context of "ubiquitous computing" and "big data"", Journal of Information, Communication & Ethics in Society, Vol. 12, No. 2, pp. 93-106.

Mcneely, C.L.; Hahm, J. (2014), "The Big (Data) Bang: Policy, Prospects, and Challenges", Review of Policy Research, Vol. 31, No. 4, pp. 304–310.

Medina, E. (2015), "Rethinking algorithmic regulation", Kybernetes, Vol. 44, No. 6/7, pp. 1005–1019.

Mohammadpourfard, M.; Doostari, M.A.; Ghoushchi, M.B.G.; Shakiba, N. (2015), "A new secure Internet voting protocol using Java Card 3 technology and Java information flow concept", Security and Communication Networks, Vol. 8, No. 2, pp. 261–283.

Montgomery, K.C. (2015), "Youth and surveillance in the Facebook era: Policy interventions and social implications", Telecommunications Policy, Vol. 39, No. 9, pp. 771-786.

Morris, T.H.; Nair, V.S.S. (2010), "Private computing on public platforms: portable application security", Wireless Communications and Mobile Computing, Vol. 10, No. 7, pp. 942–958.

Navarro, J.M.B.; Villaverde, J.C. (2014), "The future of counter--terrorism in Europe the need to be lost in the correct direction", European Journal of Futures Research, Vol. 2, No. 1, pp. 1–12.

Nickerson, D.W.; Rogers, T. (2014), "Political Campaigns and Big Data", The Journal of Economic Perspectives, Vol. 28, No. 2, pp. 51–73.

Norberg, P.A.; Horne, D.R. (2013), "Coping with information requests in marketing exchanges: an examination of pre-post affective control and behavioral coping", Journal of the Academy of Marketing Science, Vol. 42, No. 4, pp. 415–429.

Nunan, D.; Di Domenico, M.L. (2013), "Market research and the ethics of big data", International Journal of Market Research, Vol. 55, No. 4, pp. 2–13.

Ohlhausen, M.K. (2014), "Privacy challenges and opportunities: The role of the federal trade commission", Journal of Public Policy and Marketing, Vol. 33, No. 1, pp. 4–9.

Olsson, N.O.E.; Bull-Berg, H. (2015), "Use of big data in project evaluations", International Journal of Managing Projects in Business, Vol. 8, No. 3, pp. 491–512.

Ossorio, P.N. (2014), "The Ethics of Translating High-Throughput Science into Clinical Practice", Hastings Center Report, Vol. 44, No. 5, pp. 8–9.

Parham, A.G.; Mooney, J.L.; Cairney, T.D. (2015), "When BYOD Meets Big Data", Journal of Corporate Accounting & Finance, Vol. 26, No. 5, pp. 21–27.

Park, Y.J.; Skoric, M. (2015), "Personalized Ad in Your Google Glass? Wearable Technology, Hands-Off Data Collection, and New Policy Imperative", Journal of Business Ethics, p. 1–12.

Parry, I.W.H.; Walls, M.; Harrington, W. (2007), "Automobile Externalities and Policies", Journal of Economic Literature, Vol. 45, No. 2, pp. 373–399.

Phillips, K.A.; Trosman, J.R.; Kelley, R.K., et al. (2014), "Genomic sequencing: Assessing the health care system, policy, and big-data implications", Health Affairs, Vol. 33, No. 7, pp. 1246-1253.

Poole, A.H. (2014), "How has your science data grown? Digital curation and the human factor: a critical literature review", Archival Science, Vol. 15, No. 2, pp. 101–139.

Porat, A.; Strahilevitz, L.J. (2014), "Personalizing Default Rules and Disclosure with Big Data", Michigan Law Review, Vol. 112, No. 8, pp. 1417–1478.

Portmess, L.; Tower, S. (2014), "Data barns, ambient intelligence and cloud computing: the tacit epistemology and linguistic representation of Big Data", Ethics and Information Technology, Vol. 17, No. 1, pp. 1–9.

Qin, B.; Wang, L.; Wang, Y., et al. (2015), "Versatile lightweight key distribution for big data privacy in vehicular ad hoc networks", Concurrency and Computation: Practice and Experience, Vol. 28, No. 10, pp. 2920-2939.

Rastogi, N.; Gloria, M.J.K.; Hendler, J. (2015), "Security and Privacy of Performing Data Analytics in the Cloud", Journal of Information Policy, Vol. 5, pp. 129–154.

Richards, N.M.; King, J.H. (2014), "Big Data Ethics", Wake Forest Law Review, Vol. 49, No. 2, pp. 393–432.

Robbin, A.; Koball, H. (2001), "Seeking explanation in theory: Reflections on the social practices of organizations that distribute public use microdata files for research purposes", Journal of the American Society for Information Science and Technology, Vol. 52, No. 13, pp. 1169–1189.

Roski, J.; Bo-Linn, G.W.; Andrews, T.A. (2014), "Creating value in health care through big data: Opportunities and policy implications", Health Affairs, Vol. 33, No. 7, pp. 1115-1122.

Rubinstein, I.S. (2013), "Big data: The end of privacy or a new beginning?", International Data Privacy Law, Vol. 3, No. 2, pp. 74-87.

Samuels, J.G.; Mcgrath, R.J.; Fetzer, S.J., et al. (2015), "Using the Electronic Health Record in Nursing Research Challenges and Opportunities", Western Journal of Nursing Research, Vol. 37, No. 10, pp. 1284–1294.

Schadt, E.E. (2012), "The changing privacy landscape in the era of big data", Molecular Systems Biology, Vol. 8, No. 1.

Schatzmann, J.; Schäfer, R.; Eichelbaum, F. (2013), "Foresight 2.0 - Definition, overview& evaluation", European Journal of Futures Research, Vol. 1, No. 1, pp. 1–15.

Schintler, L.A.; Kulkarni, R. (2014), "Big Data for Policy Analysis: The Good, The Bad, and The Ugly", Review of Policy Research, Vol. 31, No. 4, pp. 343–348.

Schnell, R. (2014), "An efficient privacy-preserving record linkage technique for administrative data and censuses", Statistical Journal of the IAOS, Vol. 30, No. 3, pp. 263–270.

Schnell, R. (2014), "An efficient privacy-preserving record linkage technique for administrative data and censuses", Statistical Journal of the IAOS, Vol. 30, No. 3, pp. 263–270.

Selinger, E.; Hartzog, W. (2015), "Facebook's emotional contagion study and the ethical problem of co-opted identity in mediated environments where users lack control", Research Ethics.

Siemelink, A.J. Digital forensics as a service: Game on. Digital Investigation. (no prelo)

Smith, S. (2014), "Data and privacy: Now you see me; New model for data sharing; Modern governance and staticians", Significance, Vol. 11, No. 4, pp. 10–17.

Stough, R.; Mcbride, D. (2014), "Big Data and U.S. Public Policy", Review of Policy Research, Vol. 31, No. 4, pp. 339–342.

Sukumar, S.R.; Natarajan, R.; Ferrell, R.K. (2015), "Quality of Big Data in health care", International Journal of Health Care Quality Assurance, Vol. 28, No. 6, pp. 621–634.

Taylor, L.; Cowls, J.; Schroeder, R., et al. (2014), "Big Data and Positive Change in the Developing World", Policy & Internet, Vol. 6, No. 4, pp. 418–444.

Taylor, L. (2015), "No place to hide? The ethics and analytics of tracking mobility using mobile phone data", Environment and Planning D: Society and Space, Vol. 34, No. 2, pp. 319-336.

Terry, N. (2015), "Navigating the Incoherence of Big Data Reform Proposals", The Journal of Law, Medicine & Ethics, Vol. 43, No. s1, pp. 44–47.

Tse, J.; Schrader, D.E.; Ghosh, D., et al. (2015), "A bibliometric analysis of privacy and ethics in IEEE Security and Privacy", Ethics and Information Technology, Vol. 17, No. 2, pp. 153–163.

Ulltveit-Moe, N.; Oleshchuk, V. (2015), "A novel policy-driven reversible anonymisation scheme for XML-based services", Information Systems, Vol. 48, pp. 164–178.

Ulltveit-Moe, N. (2014), "A roadmap towards improving managed security services from a privacy perspective", Ethics and Information Technology, Vol. 16, No. 3, pp. 227-240.

Unsworth, K. (2014), "Questioning trust in the era of big (and small) data", Bulletin of the American Society for Information Science and Technology, Vol. 41, No. 1, pp. 12–14.

Van Den Hoven, J.; Weckert, J. (2008), Information technology, privacy, and the protection of personal data, Information Technology and Moral Philosophy, Cambridge University Press, Cambridge.

Varian, Hal R. (2014), "Beyond big data", Business Economics, Vol. 49, No. 1, pp. 27–31.

Vezyridis, P.; Timmons, S. (2015), "On the adoption of personal health records: some problematic issues for patient empowerment", Ethics and Information Technology, Vol. 17, No. 2, pp. 113–124.

Wang, X.; Liang, Q.; Mu, J., et al. (2015), "Physical layer security in wireless smart grid", Security and Communication Networks, Vol. 8, No. 14, pp. 2431–2439.

Weaver, S.D.; Gahegan, M. (2007), "Constructing, visualizing, and analyzing a digital footprint", Geographical Review, Vol. 97, No. 3, pp. 324–350.

Zhang, X.; Liu, C.; Nepal, S., et al. (2013), "SaC-FRAPP: a scalable and cost-effective framework for privacy preservation over big data on cloud", Concurrency and Computation: Practice and Experience, Vol. 25, No. 18, pp. 2561–2576.

Zhou, W.; Piramuthu, S. (2014), "Information Relevance Model of Customized Privacy for IoT', Journal of Business Ethics, Vol. 131, No. 1, pp. 19–30.

Zuboff, S. (2015), "Big other: Surveillance capitalism and the prospects of an information civilization", Journal of Information Technology, Vol. 30, No. 1, pp. 75–89.