Study on Ethics and Integrity in the Use of Big Data in Analysis and Research

Anca Ioana Iacob (Troto)
University of Craiova, Romania
anca_ioana_iacob@yahoo.com

Abstract

In a society where information is the basis of business decision-makers and its quality directly influences economic actions and activity, databases are a subject as common as it is difficult to analyze, regulate and subject to ethical norms. The present study carries out an analysis of the needs imposed by the technological evolution of recent years, in the light of the fundamental rules governing the ethics and professional integrity of the research activity. Thus, the paper seeks a common denominator between the basic principles of morality that defines the database analyst and the ethical dilemmas that arise in the stages of data processing. Through an objective analysis of the theoretical aspects, but also of the practical reality, the research seeks solutions in shaping the principles of ethics and integrity, in order to update them to the current socio-economic and academic environment.

Key words: big data, research ethics, integrity, database analyse
J.E.L. classification: C10, C40, C80

1. Introduction

In a society where information underpins decision-makers and its quality directly influences actions and socio-economic activity, databases are a subject as common as it is difficult to analyse, regulate and subject to ethical norms.

In recent years, big data analyses have become increasingly present among corporate, macroeconomists and academic analyses, with such studies significantly superior in terms of the accuracy of analyses on limited samples.

The big data concept has emerged in the IT industry and in the statistical field as a term that encompasses the use of a large amount of information or data. This opportunity was generated by technological development. We can say that the need to analyze an increasing volume of information has led to a technological leap in IT development, but on the other hand, the exponential pace at which IT resources have evolved in recent years have encouraged the use of big data, providing opportunities for unimaginable research and analysis decades ago.

Technology is changing the business world rapidly, and databases are becoming more and more part of many companies' competitive strategy. The combination of data big and high-performance analysis has the potential to provide significant organisational value, as well as support for the macroeconomic strategies of states and analyses carried out by the scientific environment. As a result, there can be no denying the need for big data in the statistical analyses of the present and the future, but for optimal results their use involves rules, a good-performing human resource, good practices and objectivity.

Big data can be considered a method of approaching research that addresses a complex analysis, systematic minating of information or handling data series that are too bulky, heterogeneous, too complex to be analyzed by manual processing or using classical software. The purpose of using big data is primarily to generate greater statistical power, for broad conclusions, it exposes new complex and relevant statistical approaches or results that go beyond the analytical capabilities of the human brain.

In this context, the principles of ethics and integrity governing statistical research, both for business and for the scientific environment, are the subject of this work, with the aim of harmonising the fundamental principles of morality with the rapid evolution of technology.

2. Theoretical background

The concept of big data entered the literature in the 90s, broadly defining the concept in a similar way. The literature includes a number of definitions, depending on the source of citation and the scope of the literature in question, but they are fundamentally similar. Gartner IT Glossary defines big data as " high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation." A similar definition is set out by the TechAmerica Foundation in 2012: "Big data is a term that describes a large volume of high-speed, complex and variable data that requires advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information."

The term big data refers to the extraction, management and analysis of data sets that are too large to be treated in a regular way. As a result, research and analysis involving big data can only be carried out using special software and, in many cases, especially dedicated hardware. Kaplan and Haenlein (2019) define big data as "data sets characterized by huge amounts (volume) of frequently updated data (speed) in various formats, such as numeric, textual or images/videos (variety)". Another defining description is issued by De Mauro, Greco and Grimaldi in 2016, describing big data as " the information asset characterized by a high volume, speed and variety, which requires specific technology and analytical methods for its transformation into value". Mihet and Philippon (2019) argue that the big data and artificial intelligence technologies that their processing involves can be viewed from three perspectives: as an intangible asset, as a search and match technology and as a forecasting method.

If three decades ago it was a mathematical concept, a pioneer in academic circles and an aspiration of researchers in the field of statistics, today the big data and the complex processes it involves is a constant concern of the authorities from a legislative point of view. Given the complexity, and in particular the implications of the use of large-scale data, the authorities are transposing big data into a topic of interest, constantly updated by the legislative for a. With the aim of regulating, avoiding abuses and respecting the fundamental rights of the person (whether physical or legal), specific legislation has arisen at local and international level, which delimits and defines the acceptance of data processing and, in particular, the use of the results arising from such analytical studies. At European Union level there is no mandatory definition but, according to Opinion 3/2013 of the European Data Protection Working Group, "Big data is a term that refers to the enormous increase in access and automatic use of information: it refers to the huge amounts of digital data controlled by companies, authorities and other large organisations, which are subject to extensive analysis based on the use of algorithms. Big data can be used to identify general trends and correlations, but can also be used to directly affect people."

The use of a large amount of data is applicable in three directions:

- Integrating information by concatenation from several dispersed sources and transposing it into a common format through processing into a unitary form.
- Manage and store data on an electronic media, either on-premises or cloud, while meeting the requirements of fairness and confidentiality.
- Analyse the data to obtain a clear view of the varied data and explore them for new conclusions.

Big data is in a permanent change, applicable in all industries. Finance is becoming one of the most promising areas of management and governance that finds applicability to data analysis. Big data significantly changes business models in financial companies. Many researchers argue that big data fuels the transformation of finance and business in general in ways and perspectives that we cannot yet fully assess (Sun, Zshi and Zhang, 2019).

In the financial field, big data represents large sets of complex information that, following an analysis, can lead to an increase in business intelligence by identifying trends and determining forecasts based on statistical methods, improving business operations and decision-making

processes, determining economic and financial developments and the list of applicability can continue. Once introduced into statistical practices and scientific research, the term big data gives rise to a number of academic concerns regarding the content of the data, its relevance, data extraction, storage, data analysis, search method, classification and archiving, transfer of information, its presentation, querying, updating, sources of information and, last but not least, the demarcations of professional ethics and integrity that the use of big data implies.

With regard to the defining features of big data, the literature presents common opinions, noting over time a detail of the definition of "V" characteristics. Chen and Zang define this subject in 2014 in a simplified form, finding three defining features: volume, speed and variety. In 2016, Abbasi developed this concept, mentioning four characteristics, adding to those mentioned , the veracity. A more complex approach is the one supported by Seddon and Currie (2017), which issue the theory of the seven "V"s: volume, speed, variety, veracity, variability, value and visualization.

The view issued by Seddon and Currie is confirmed in the literature and by other authors, one example being Alexander and Coardos (2017), who, however, attribute more importance to four of them: volume, speed, variety and veracity. Starting from the intrinsic definition of big data, the fundamental characteristic is the volume of information. Based on the fundamental statistical theories, the approach of a research becomes all the more developing and generates results all the more representative the more voluminous the data series that represent the basis of calculation. On the other hand, however, too much data can lead to storage problems, as well as the logistical and technological capabilities available to the research entity. It is precisely for this reason that in research, government and business media there is the practice of extracting and storing a huge amount of operational, public, commercial or social data archived for "hibernation", anticipating future opportunities to increase the speed and processing capabilities of data, which will lead to new conclusions and estimates, currently inaccessible from a technological point of view.

A permanent challenge in the use of big data is the dilemma that scientific research has when deciding to extract and filter information, i.e., whether it is necessary to use complete data to draw certain conclusions about their properties or a sample is sufficient to generate similar results, minimising effort, simplifying resources and, by implication, reducing costs. The big data contains the axiomatic dimension-related trait. But statistical sampling allows the filtering of accurate data collection points from a wider set to estimate features of the entire set of elements. Big data can be filtered on different categories of data in the sample selection process using sampling algorithms specific to big data statistical research.

The speed at which information is extracted, classified, managed, transferred and analysed is another determining feature. This refers both to the speed with which the data are produced and to the speed at which the data must be processed to meet demand. This involves data streams, recording structuring, and the technical possibility of accessing and delivering the beneficiary. The speed of data operation is constantly developing, thanks to the continuous technological development, minimizing the physical volume of information storage and the advantage of optimal data distribution and classification, the continuous development of new methodologies and software, the digitization of the scientific environment and easy access to information beyond geographical barriers. The high speed of data processing must evolve, in the context where decision-making capacity is based on this information and receiving results in a short time may be essential for business.

The variety of information is also an important feature, giving the researcher the opportunity to process the data and express and materialize the results by easy methods to understand. The variety refers to the plurality of structured and unstructured data sources, which, inter alia, include different sources on graphics, audio, photo, video and text media (Constantiou and Kallinikos in 2015; George and others in 2016). In scientific research, it is not sufficient for databases to be bulky in order to generate relevant results. The variety of information is essential for data processing. The use of homogeneous data series exposes scientific analysis to the risk of irrelevant conclusions, and in the case of predictions, to results that are not applicable or not unanimously accepted. In order to avoid such inadequate analyses, it is still necessary to give increased attention to the quality of the sample used and to the filters to which the data are subjected in the pre-processing activity. Due to the degree of technology and globalization, this desire is now much more accessible, the variety of information sources including tabular data (databases), static images, metering data, financial data, graphic

developments, hierarchical data, documents, specialized software, electronic messages, press information, etc.

The veracity of the data refers to the degree of trust or distrust that the processed data inspires. Big data quality is less controllable because it comes from different sources, presentation form and complexity that is difficult to expose to end-users of this data. The ability to assess the compliance, accuracy and honesty of the data under analysis is essential for the research environment. Here the discussion is carried out around the responsibility of the initial data generator, the purpose for which the data is issued and the reactions of the receivers, when the latter may show mistrust caused by the inevitable opacity in the process of data extraction and processing.

3. Research methodology

In the above presented, through a bibliographical information and documentation were shown the theoretical-conceptual aspects of big data practice in the business and contemporary academic environment, in the light of the features that define the concept of processing with a large volume of data, but also in terms of ethical conduct and professional conduct. The use of a large amount of data involves rules and practices whose compliance is indispensable for achieving valid and credible results and conclusions.

For a detailed analysis of the benefits and challenges that the use of big data imposes on statistical activities and for drawing conclusions on the harmonisation of fundamental principles of ethics and integrity with the rapid development of technology.

The collection of this information, combining the analysis of the functionalities of the big data with their theoretical aspects, and in particular with the attributes presented in the form of the characteristics "V", results the main principles of ethics and integrity that big data imposes on scientific research, applicable in the existing technological perspectives and limits, but also certain challenges related to professional morality and whose definition evolves in parallel with the new methods and possibilities that technology makes available to statisticians.

4. Findings

The science is based on fundamental principles, valid at international level and applicable in all scientific fields. A first fundamental principle is honesty with oneself and those around you. Honesty is an ethical principle that underpins all the rules and adapts to a discipline function, as well as to the technological and scientific evolution of the sector they treat. The role of scientific research is to transmit information, to perpetuate knowledge of humanity, but first of all to assess and advance knowledge through the prism of fairness and good faith.

In the paper published in 2016 on the rules of conception, drafting and publication of scientific materials, Dinu, Săvoiu and Dabija consider that the issuance of truth in a coherent and validated form represents the main purpose of any scientific research process, and the achievement of this desire can only be achieved by following rules, a standardized procedure and rules unanimously accepted by academia.

Ethics in the field of big data, found in the literature and under the name "Data Ethics", consists of an interest in the systematization, defence and recommendation of concepts of behaviour and good practice in terms of data extraction, filtering and processing, as well as the exposure of results. The ethics of big data has an important component with regard to personal data, but this paper deepens the subject through the prism of big data ethics used in the case of scientific research in the field of finance. In this case, the confidentiality of personal data and the personal consent of the individual are less subject to the investigation and are considered irrelevant. As a result, big data ethics is thorough in terms of information used in business, financial and econometric studies.

The ethics and integrity of data extraction and processing involve several principles underlying the research and without which the results of the analyses would have no credibility:

- the principle of ownership, refers to the owner of the information and respect for his rights;
- transparency in the data processing process, the owner having the right to inform on how to collect and process the data, its use and the supporting platforms for the reproduction of the results;

- confidentiality, where the data involved are not of a public nature;
- accessibility to processing methods.

Big data ethics in the financial and statistical fields raise the academic field with a number of issues that conclude rules of good practice that support scientific research in a context of fairness and objectivity. The most important aspects of the ethics and integrity of big data use are: consent, transparency, ownership, digital equality, trust.

Consent is a basic principle with an impact in the first phase of statistical research. It should be noted that in the area of macroeconomic and financial research, statistical information is often transparent, its publication being an obligation on the part of the processor. Here we illustrate with the financial data that are required to make public a public listed company or with the macroeconomic indicators of the European Union states, which are the source of analysis for the statistical journals of the European Commission. The ethical question that the principle of consent raises is that of long-term agreement and of ensuring the possibility of withdrawing information from the public environment if, after acceptance, the owner of the data decides its confidentiality.

The transparency of statistical information refers to the data owner's right to track how the information he provides is processed, analysed and published.

Another issue that the ethics of data processing raise is that of their ownership. Fundamentally, the concept of ownership implies exclusive ownership and, by implication, the impossibility of owning another entity without the prior consent of the owner. The rules on ownership of financial information are constantly changing, representing a current concern of the authorities, who find legal solutions to solve the ethical dilemma.

Digital inequality is regarded by researchers as one of the problems that the use of big data raises. This refers to the lack of fairness arising from the fact that few entities have access to the infrastructure and logistics necessary to access a large volume of data, so statistical analysis is only available to certain processors and thus generates the dilemma of fairness and subjectivity with which the final results are interpreted.

Confidence in analytics using massive data series may also be questionable, if the complex algorithms they involve in the analysis are difficult for users to understand. This ethical issue is resolved in the use of data from credible sources only, official and by mentioning the used sources.

In the process of conducting scientific research, it is necessary to respect the ethical and integrity principles in all phases of manipulation of the big data. Thus, in the first phase, that of accessing the data, it is necessary to fulfil the conditions of acceptance, in accordance with the principle of consent which the owner of the data provides. Access to data is subject to the regulations of professional ethics from the very beginning of their reading. Note that by simply accessing the database, the researcher needs the consent of the information, even if this information will not be used or replicated. Therefore, before accessing, it is important to analyse the source and provenance of the information in order to avoid a breach of confidentiality, but also to ensure the veracity of the information.

In the second phase of data operation, that of data collection, it is particularly important to format the data, obtaining it in standardised models and comparable formats, especially in the case of comparative research. Note that depending on the particularities of scientific research, the information collected should comply with certain specific standards or reference systems. Due to technological advances, data collection methods are not constrained by technological constraints, but rather by the researcher's capabilities in terms of their use. This is where the latter's ability to give the study an out of the box perspective comes in. Modern technologies also allow data collection by less common methods, such as perception sensors, web traffic monitoring or automatic extraction of information from online platforms. These collection methods have developed in recent years with a particular scale, so that academia and legislatures are having difficulties and are hardly keeping up with them, the definition of ethical principles being constantly updated. On the basis of these principles themselves, lawsuits have been filed against companies engaged in web data mining activities (Vaughan, 2013), misused in commercial actions or for the purpose of manipulation. In case of doubt as to the legality of collecting data from a website for academic use, it is best for the author to contact the information provider directly to obtain explicit consent. This is where a new challenge arises that the researcher may face, namely that the platform he is accessing for data collection is an intermediary of information, which is why its acceptance may be insufficient to

comply with security and ethics protocols; in this case, it is mandatory for the collector to ensure that the original source of the information and the rightful owner of the data are entitled.

In terms of storing and archiving data, often the storage capacity required exceeds that found in personal computers. The way data is stored raises on the one hand the question of logistics and technological resources, and on the other hand the question of data security. How data is stored depends on its volume. For relatively simple datasets, no dedicated storage solutions are required, with software such as Excel, SAS, or SPSS able to handle variable information content. For example, Excel may contain 1,048,576 rows and 16,384 columns (Microsoft, 2016). However, there is a situation where these programmes are not sufficient to store the big data needed for research, and a more complex approach is needed. A structured database, such as SQL, can be considered. Solutions with parallel tubulars, open-source relational databases such as MySQL and PostgreSQL are also available. In such situations, the controller must have specific it knowledge in order to be able to store and subsequently process the data. Another storage option is in the form of a cloud, with the advantage of personalized costs over user needs, with the storage fee depending on the amount of data held. The structured and distributed storage solution is also to be mentioned, exemplified by the Apache Hadoop solution, which allows data allocation on multiple computers. It should be noted that this solution requires knowledge of it and DBA infrastructure (data base administration), the access of which can significantly supplement storage costs. From the point of view of data storage ethics, greater attention is needed to monitor persons who have access to big data series and to comply with the confidentiality conditions concerning specialists who have access to them for the purpose of IT maintenance (where appropriate requiring the signing of confidentiality clauses and the provision of security, monitoring and access protocols). Cases known as 'information leak' are not few, and in many cases, this may jeopardise the final findings of the study, the prohibition of the use of information or academic compromise of the authors of the research.

In the next phase, i.e. data processing, an important action is to filter the big data, with the aim of obtaining structured and representative information. This representativeness of the information is a challenge for the author of a scientific study, and any errors that arise during this information processing phase may jeopardise the relevance of the data analysed. In order to retain the principle of integrity, the database operator should impose clear, rational and neutral rules on data filtering. It should be recognized that researchers can often be emotionally involved in the study they undertake and may irrationally decide to process data subjectively. In such cases, professional ethics requires recognition of this situation on the part of the author and we propose in this respect the involvement of a third party, which proves neutrality and objectivity. This may include a collaborative researcher or the use of consulting services (but do not forget in this case that such academic affiliation must comply with the principles of confidentiality and integrity of all study participants).

Data analysis is the most important phase of the process of using big data, and also the one that imposes professional ethical rigours. Unfortunately, the more detailed and specific these rules and standards are, the greater the likelihood of becoming contradictory and inconsistent (M. Popa, 2016). Data analysis is a process of inspection, filtering, processing and data transformation to issue new discoveries, for academic purposes or for the purpose of being used by a beneficiary. Data analysis has several approaches, including various data discovery techniques and methods, descriptive analysis, reports, tables and graphic views, predictive analysis through statistical models to better understand future behaviour, and prescriptive analyses that clearly suggest solutions.

From the perspective of professional ethics, data analysis must be primarily a quality one. Even if we are talking about big data, so quantity, a large volume of processing, when the researcher reaches the data analysis phase, the volume of information is transposed into quality. In the view of Ciora and Buligiu (2013), data quality analysis involves five steps:

- Setting the objectives of data quality analysis and the structure of the information collected: is carried out before the evaluation of the information in order to structure and filter the data in the database subject to the analysis;
- Verification of preliminary data: validation of the form of quality assessment reports, completion of basic statistical calculations and preparation of tables and graphs.
- Selection of the statistical test: the optimal methodology for the calculation and analysis of the data will be selected, based on the previous stages of the research. At this stage, the researcher identifies the key parameters needed to validate statistical procedures;

- Verification of forecasts: assessment of key parameters or preliminary data, if they are characterised by acceptable values;
- Issuing of the results: the calculations necessary for the statistical test and the documentation from which the inferences of these calculations are derived.

The ethics of big data analysis is a constant concern of academia. However, ethical dilemmas are often viewed superficially. Analytical work involves respecting the research and validation methodologies for the accuracy of the resulting data, in order to always give due attention to ethical issues.

The ethics of big data analysis involve not only the need for the accuracy of the research methodology, but also the accessibility to information of those concerned, as well as the manipulation of information through the prism of good will. The accuracy of the processes refers to the observance of the research phases, the choice of the appropriate methods of analysis, and the clear definition of the purpose and destination of the final results. In order to be effective and correctly applied, it is necessary that the methodology used in the analysis of data be able to achieve the objectives of scientific research, be effective, highlight the essential points of research and answer the questions set out in the hypothesis. By respecting these ethical principles concerning the methodology of the research study, one can also avoid the cognitive predisposition of the researcher for the involvement of stereotypes or personal and subjective patterns on the ongoing analysis. By analysing data without objectivity and integrity, research can be compromised by inadequate conclusions, ignoring results or issuing preconceived opinions.

It should be noted that the analyst is prone to human error, and the principles of ethics and integrity are intended to avoid these human errors as much as possible. Thus, it should be noted that professional ethics and integrity can be violated by the researcher by premature opinions, by inappropriate analogies, by lack of empathy or, on the contrary, by the abusive involvement of emotions in the professional act, the denial of reasoning, the intentional disregard of certain information, the exaggerated confidence in pessimistic or optimistic prejudices, ignorance in the sense of insufficient knowledge of the context of analysis.

The last phase of the big data study is the reporting and visualization of the results. This step is intended to facilitate interpretation by the auditor concerned or by future users of the results of the scientific study. A first challenge for the data analyst is to represent the conclusions in accordance with the ethical principles of transparency, in a way accessible to the non-scientific environment, especially in the case of those analyses using big data and complex methodologies, which are often not tangible to the general public. In this respect, professional ethics requires the scientific researcher to present the conclusions of the study by mentioning the source of the information, the methodologies used and representing the conclusions as clearly and concisely as possible.

Rosnow and Rosenthal define in 2011 ethics and integrity criteria that reporting research data must meet, concluding a number of information characteristics. This must be:

- transparent – the results are presented openly, respecting honesty and through an appropriate platform and technical language, within the auditor's reach.
- informative – contain enough information so that the auditor correctly perceives what the researcher wants to convey;
- accurate – reported results with the accuracy required by the situation under consideration;
- correct – avoiding and correcting errors in measuring, calculating and reporting numbers, but also adopting a neutral tone, free of exaggerations and subjectivity in issuing conclusions;
- substantiated – the research methodology should be chosen correctly; the statistical calculation methods should be appropriate to the data and the assumptions issued should be verified before being entered in the study conclusions.

Another issue that professional integrity in academia raises is conflict of interest. In order to preserve ethical and professional principles, the analyst who operates big data for the purpose of research has a moral and legal obligation not to distort the analysis process due to personal interests. A manifestly emotional involvement or interested in a purpose other than that specific to the research undertaken leads unequivocally to the compromise of the entire study. The ethical principle of trust in the objectivity and professionalism of the researcher who draws conclusions is fundamental for academia. The same register also includes the incompatibility of carrying out a statistical analysis

using data which contradicts the position, function or interests of the scientific researcher. And in this case, the study conducted is jeopardised and the results of the analysis are irrelevant.

Ethics in academia and, in particular, ethics in the digital environment impose rules on dignity. "Human dignity is inviolable. It must be respected and protected. "Technology should not dictate values and rights, but neither should the relationship between them be reduced to a false dicotomy," says the Charter of Fundamental Rights of the European Union. The importance of big data use, cloud storage technologies, high speed of information propagation, Internet opportunities – all of these are recognised, and the benefits they bring to economic development, international cooperation and the business and academia are self-evident. However, the ethical side of this development, which cannot be limited to the legal framework, but must be understood from a human, moral and philosophical point of view, must not be ignored. At European Union level, the European Data Protection Supervisor (EDPA) is launching the concept of a 'large data protection ecosystem', identifying four methods to respond to the challenges that the exponential development of digitalisation manifests:

- the implementation of a forward-looking legislative framework that contributes considerably to the ethical principles of data processing;
- professional and responsible data controllers who determine the processing of information including personal data;
- the creation of data processing products and services that take into account privacy;
- more rights for natural persons.

Avoiding plagiarism remains one of the topics of interest in recognizing the work of a researcher in academia. Analysis of high-volume data and information is not immune from this error that an analyst, willful or by mistake, can make. Applying on the big data field the methods mentioned by Stephen (2018) to avoid plagiarism, we can consider that it is absolutely necessary for the researcher to carry out his work on the basis of honesty, recognizing the paternity of the ideas presented, knowing the ways of quoting sources of inspiration and always mentioning in his works the source of the data processed and used in the research methodology, in order to draw conclusions.

Without developing the obviously negative effects of plagiarism, it is worth mentioning that technology and the very collection of a large volume of specialized information, software and platforms are tools by which the scientific researcher to avoid plagiarism. It should also be said that plagiarism can occur in a statistical study from the data collection stage, which is why mentioning the sources of export of big data is imperative. It is also necessary that the analysis of information, the issuing of conclusions and the visuality of reporting be subject to creativity, personal input and innovation.

Globalisation is another factor that interposes in the ethical-scientific research relationship and is all the more visible in the field of big data, due to the specific internationalisation of this field. In this way, we see a trend of common global interest in ethical issues in science, internationalisation which at European level has resulted in the emergence of consensual recommendations, the issuance of the European Researcher's Book and the Code of Conduct for the Recruitment of Researchers, but also through cooperation at institutional level, namely the establishment of National Councils and Ethics Committees. The European Group on Ethics in Science and New Technologies (EGE) was also established as a neutral, independent, pluralistic and multidisciplinary body, which has the explicit role of engaging in the regulation of the technological field and advising the European Commission on the ethical aspects of science and new technologies and underpinning the preparation and preparation of European legislation and policies.

Finally, we are in agreement with Emilia Şercan, who mentions in her paper "Academic Deontology – Practical Guide" (2017) that the lack of academic integrity has negative, long-term effects on society. Applying on the field of big data, we can say that the lack of professional ethics and integrity in research can lead to the impairment of the quality of the education and specialization process in the field of statistics, the invalidation of academia by public opinion, the underappreciation of teachers and courses, the stigmatization of scientific research and the devaluation of high-level studies. To all this, we add long-term consequences for the quality and volume of the research process and, by implication, theoretical and practical innovation.

5. Conclusions

Big data analysis in scientific studies is subject to the rules and regulations of ethics and academic integrity, but having specific peculiarities in the statistical field. The exponential development that the big data field has been experiencing in recent years makes academic and legislative environments have to support and continuously improve ethical principles and rules. Through an analysis of the particularities of this field and observation of strengths and weaknesses, as well as future opportunities and challenges, it was possible to identify professional ethics and integrity issues that are required of scientific researchers who use a large amount of information in their studies.

An analysis of the literature, as well as the legislative framework implemented at European Union level, notes that the ethics of big data analysis have common points with the ethical principles governing the statistical field, but also have traits whose definition and regulation are necessary to be improved with technological evolution.

With the emergence of new research methods, new sources of information and data processing methodologies, the principles of integrity and ethics feel the need for evolution. International organisations, as well as academic institutions, are designed to foster dialogue between academia, authorities and the general public, to provide advice on ethical and integrity issues, to instil principles of scientific ethics in both future generations of researchers and public opinion, to ensure an intellectual environment with the aim of changing ideas and issuing recommendations.

We can say that morality is changing, updating to society, technology and economical changes, but the fundamental principles of fairness, integrity, objectivity and professionalism that define human civilization remain the same. It is only a model of the principles of ethics, an update of them to cover new needs arising in the field of research, but all for the purpose of extrapolation and refining, not to modify the concept of "what is right".

6. References

- Abbasi, A, Sarker, S., Chiang R.H., 2016. Big data research în information systems: toward an inclusive research agenda. *Journal Association Inf Syst* 17(2), pp. 1–32
- Alexandru, A., Coardoș, D., 2017. Big Data - Concepte, arhitecturi și tehnologii [Big – Data. Concepts, architectures and technologies]. *Romanian Review of Informatics and Automatics,* vol. 27, no. 1
- Chen, C.P., Zhang, C.Y., 2014. Data-intensive applications, challenges, techniques and technologies: a survey on Big Data. *Inf Sci* 275, pp. 314–347
- Ciora, L., Buligiu, I., 2003. Metode și tehnici de analiza a calitatii datelor [Methods and techniques for data quality analysis]. *Economic Informatics Review,* no. 1(25)
- Council of Europe, European Union Agency for Fudamental Rights, 2014. *Handbook of European data protection legislation*
- Constantiou I.D., Kallinikos, J., 2015. New games, new rules: big data and the changing context of strategy. *J Inf Technol*, vol. 30(1), pp. 44–57
- De Mauro, A., Greco, M., Grimaldi, M., 2016. A formal definition of Big Data based on its essential features. *Library Review,* Vol. 65 No. 3, pp. 122-135
- Dinu, V., Săvoiu, G., Dabija, D.C., 2016. *Design, write and publish a scientific article*. Bucharest: ASE Publishing House
- European Data Protection Supervisor, 2015. *Opinion 4/2015 - Towards a new digital ethic. Data, dignity and technology*Gandomi, A., Haider, M., 2015, *Beyond the hype: Big data concepts, methods, and analytics,* International Journal of Information Management, no. 35, pp. 137-144
- Gartner IT Glossary; http://www.gartner.com/it-glossary/bigdata/
- George, G, Osinga C.E., Lavie, D., Scott, B., 2016. Big data and data science methods for management research. *Academy Management Journal*, vol. 59(5), pp. 1493–1507
- Haenlein, M., Kaplan, A., 2019, A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, vol. 61/4
- Iovan, S., Daian G.I., 2012. New Challenges: Big Data and Consumer Intelligence. *Annals of Constantin Brâncuşi University of Târgu Jiu, Engineering Series*, no. 4
- Lawson, R., Hatch, T., Desroches , D., 2019. *The Impact of Big Data on Finance Now and în the Future,* Institute of Management Accountants

- Mihet, R., Philippon, T., 2019. The Economics of Big Data and Artificial Intelligence, Disruptive Innovation în Business and Finance în the Digital World. *International Finance Review,* vol. 20, pp. 29–43
- Ministry of Education and Research, National Council for Ethics of Scientific Research, Technological Development and Innovation, 2020. *Guide for Integrity in Scientific Research*, www.research.gov.ro
- National Council for Ethics of Scientific Research, Technological Development and Innovation (CNECSDTI), 2019. *National Code of Ethics for Scientific Research*
- Nethravathi, P.S.R., Bai, G.V., Spulbar, C., Suhan, M., Birau, R., Calugaru, T., Hawaldar, I.T. & Ejaz, A., 2020. Business intelligence appraisal based on customer behaviour profile by using hobby based opinion mining in India: a case study. *Economic Research-Ekonomska Istraživanja,* 33(1), 1889-1908, DOI: 10.1080/1331677X.2020.1763822.
- Oracle.com, *What is big data.* [online] Available at: www.oracle.com/ro/big-data/what-is-big-data.html
- Popa, M., 2016. *Research methodology (course notes): 12 Criteria for evaluating scientific articles and research ethics.* [online] Available at: http://www.apio.ro/ upload/mc12_eval_etica_cercet.pdf
- Rosnow, R. L., & Rosenthal, R., 2011. "Ethical Principles în Data Analysis: An Overview" in A. T. Panter & S. K. Sterba (Eds.), *Handbook of ethics în quantitative methodology.* New York: Routledge
- Seddon, J.J., Currie, W.L., 2017. A model for unpacking big data analytics în high-frequency trading. *Journal of Business Research*, vol. 70, pp. 300–307
- Sfetcu, N., 2019. *Etica Big Data în Cercetare [Big Data Ethics in Research].* Bucharest: MultiMedia Publishing
- Sun, Y., Shi, Y., Zhang, Z., 2019. Finance Big Data: Management, Analysis, and Applications. *International Journal of Electronic Commerce,* vol. 23:1, pp. 9-11
- Șercan, E., 2017. *Deontologie academică. Ghid practic [Academic deonthology].* Bucharest: University Publishing House
- Ștefan, E.E., 2018. *Etică și Integritate academică: curs universitar [Ethics and academic integrity: university course].* Craiova: Pro Universitaria Publishing House
- TechAmerica Foundation's Federal Big Data Commission, 2012. *Demystifying big data: A practical guide to transforming the business of Government*. [online] Available at: www.techamerica.org
- The European Commission's High Level Expert Group on Artificial Intelligence 2019. *Ethical Guidelines for Reliable Artificial Intelligence (AI)*
- Vaughan, B. 2013*, AP, news aggregator Meltwater end copyright dispute*, Reuters, www.reuters.com