8-8-2022

Official Statistics: Historical Trajectory and Contemporary Threat Related to Big Data Adoption

Fabio dos Santos Cardoso Mr

Official Statistics: Historical Trajectory and Contemporary Threat Related to Big Data Adoption

Completed Research Paper

Fabio dos Santos Cardoso
University of Minho
id7039@alunos.uminho.pt

Abstract

Official Statistics are the relevant set of information to support decisions in modern societies. This relevance is a result of a historical process summarized in this Completed Research Paper. Based on this historical trajectory, the study advances to the current features of this business called Official Statistics. Information life cycle and the Business architecture, defined by Radermacher (2020), are discussed and checked under the perspective of Big Data adoption. In sequence, the features of the Official Statistics and their operators, the National Statistics Offices, are analyzed considering the risk of a failure in the Big Data adoption process. The methodology mixed literature review with documental research. Recommendations for future studies are proposed in the end.

Keywords

Official Statistics, National Statistics Offices, Big Data, Adoption

Introduction

The Official Statistics (OS) represents a huge and worldwide information system (IS). A system integrated by Official Organizations located in many countries and geographic regions. Across the centuries, Official Statistics have supplied governments and societies with demographic, economic and social data. The data is the raw material of the Official Statistics. In this context, the collection, processing, and communication/dissemination of the Official data represent the core operation of the Official Statistics agents. In view of the subject's relevance, this Completed Research Paper's objectives are to descript the historical evolution of the OS, explain its business cycle, and the challenges for the OS operators to adopt a current technology, Big Data.

There is a justification for this research: if data is the raw material for OS, the OS is the raw material to support decision-making processes in modern societies. Governments, at local, regional, or national levels use the OS to develop, deploy and evaluate public policies. Companies and businesses base their strategies and plans on economic data from OS. Citizens realize worsening or improvement in their well-being when looking at an inflation index. So, the importance of OS to modern societies is extraordinarily strong. If OS become technologically obsolete, the nations can suffer from the lack of updating. To address this relevant topic, the following paragraphs and sections discuss the contemporary threat related to Big Data Adoption by OS. In face of this reality, it is necessary to start with an explanation of the central concepts involving OS.

Firstly, it is necessary to define and differentiate the terms related to Official Statistics. Radermacher (2020) brings definitions to three main concepts to Official Statistics field: (i) statistics, (ii) statistical results and (iii) statistical institutions.

Radermacher (2020, p.1) defines statistics as "*the science of learning form data.*" According to the author, this science produces statistical results that "*are used for all conceivable information and decision-making processes*". In the same path, Radermacher (2020, p.2) remarks that "*the producers of statistics*" are statistical institutions.

Alias, the produce and use of statistics can be understood as daily-basis routines into a company, a non-governmental organization, and a soccer team. On the other hand, statistics can be the core information of an official report generated by a governmental authority. This second possibility refers to the Official Statistics. Official Statistics means "*any set of statistics produced by an organization named under secondary legislation and described by that organization as an official statistic or part of a set of official statistics.*" (UK Statistics Authority, 2022)

Those mentioned organizations are the producers of statistics as highlighted by Radermacher (2020) who also called them as Statistical offices. The author categorizes those organizations by geographic coverage level, from international, regional or local level. In the three categories, those organizations are part of or associated with the national public administration.

Statistical or statistics offices are the Official Statistics producer. This type of statistics can be differentiated by mainly two general groups, according to the UK Statistics Authority (2022): (a) "*National Statistics, which have been assessed by the statistics office as fully compliant with the regulation*"; (b) "*experimental statistics, which are newly developed or innovative statistics*". About the mentioned regulation, it refers to a set of rules to guide the planning, collection, analysis, and dissemination of the data processed by the statistics offices. Those rules can be laws, acts, or administrative codes to ensure that the Official Statistics have been produced with ethics and high standards.

As mentioned before, the Statistical Offices are public organizations working in three different geographic levels. In this research are considered the organizations which work in national and international level. This distinction is necessary to highlight the cooperation network between Statistical Offices.

Those initial features indicate that Official Statistics and its operators are not a Twenty-first Century's phenomenon. All those statistics office, their outputs and outcomes, the methodologies and techniques are results from a large historical trajectory. This trajectory of adjudicative knowledge and constant innovation is described in the next section.

Finally, a question emerges in this context: in the current days, how can an average citizen be impacted/benefited by Official Statistics with the use of Big Data? This is a huge challenge to be addressed by the Statistical Offices. The answer for this research question is proposed in this study.

About methodology, it was adopted a mix of methods, using literature review associated with documental research (Cresswell and Cresswell, 2018) to identify the historical path, and the business features involving OS. Related to data collection, the study focused on the reality of three National Statistics Offices: Instituto Nacional de Estatísticas (Portugal), Instituto Brasileiro de Geografia e Estatísticas (Brazil), and Office for National Statistics (United Kingdom). Documental research, including the content of those NSOs' websites were studied under the perspective of the content analyzes (Bardin, 2016; Van Campenhoudt et al, 2017).

A summary of the Official Statistics trajectory

Kotz (2005) and Stigler (1986) highlight the evolution of the statistics as a science and the Official Statistics as a subfield of this science. This evolution can be considered part of the European Modern History. However, those authors diverge about the time line. Kotz (2005) adopts a longer trajectory, since the Sixteenth Century.

Based on those historical references, the use of counts and number to measure military capabilities and taxations and remotes before the Common Age (Kotz 2005) and persisted during the Middle Age (Instituto

Nacional de Estatísticas, 2022). In Sixteenth Century, seminal studies and publishing took place in three European countries: German, United Kingdom, and France.

As remarked by Kotz (2005), during the 17th Century, in German, the publishing from Johann Peter Süssmilch (1707-1767) represented the initial event to the Official Statistics in that European country. As a Protestant Pastor, Süssmilch recorded the religious events, such as marriages, baptisms, and funerals. Indeed, those registers generated the initial tables and data about immigration, emigration, nationality, age of marriage, and birth rates. Many years later, in the 19th Century, the Royal Prussian Statistical Bureau adopted Süssmilch's methods in its processes to produce Official Statistics.

Following the Kotz's (2005) time line, the United Kingdon (UK) represents other zone were the statistics and Official Statistics methods rose early. The author remarks the word "statistics" is an appropriation from the Germany similar "statistik". John Gaunt (1620-1674) was one of the first involved in social and demographic measurements in UK. As a London citizen, he published a book with the metrics of the mortality in that city, based on official registers.

Another relevant British author for statistics and Official Statistics was John Sinclair (1754-1835). Kotz (2005) remarks Sinclair studies about statistical accounts in Scotland adopted an innovative method of data collection: the questionarie. The compliation from questionaries aplied in 938 Scotish Parishes and reponded by the respective ministers generated the data explored by Sinclair. That was a social statitical survey depploied over than two hundred years ago.

Still about UK's Official Statistics evolution, during the first half of the 19th century, the Royal Statistical Society was founded in London (Royal Statistical Society, 2022). This intitution, with one hundred eigthy-eigth years of history, is the active think tank for Official Statistics in the British lands.

The third country remarked by Kotz (2005) is France. The oldest register of the Official Statistics in France is a public finantial publishing from 1581. This historical register was followed by surveys about administrative data in the second half of the Sixteenth Century. In fact, the French focus to Official Statistics crossed the centuries, and until the French Revolution, the progressed kept the path. Chaptal, Neufchateau, Laplace and Duvillare are some names of intelectuals with contributions to Official Statistics development in 18th and 19th Centuries.

In the same period, Adolphe Quetelet, a Belgic scientist, took place in the French scene. Its studies were concentrated in demographic themes. His main concerne was about to adress three questions: (i) *"what are the laws of human reproduction, growth, and physical force" that influence a country or society?"* (ii) *"what influence has nature over man?"*, (iii) *"can human forces compromise the stability of the social system?"* (Quetelet, 2013).

As result of the attempting to answer those questions, Quetelet developed a theroritical construct still used until nowadays in Official Statistics, the average man. The proposal was to achive the possibility to understand a great set of data, such as a society. So, Quetelet understood the citizens as units of this society. Attempting to reduce the extremes in matters as behaviors, physical features and natural actions, he found a statistic tool: normalize the society features in the average, an average man. This concept was created to concentrate in a symbolic-model-citizen, predominant elements from each society. He pointed his intention was not to create a flat view of mankind. France should have its own average man, such as England, German and Scotland should have their own.

Until the Twenty-first Century, Quetelet's (2013) legacy still works. In 2010, Office for National Statistics[1] did a research to define who was the average man and the average woman in England (BBC, 2019).

Moreover, the Official Statistics relevance had register in other countries during the 19th Century, such as Portugal and Brazil. In 1864, Portugal ran its first demographic Census, adopting modern statistical methodologies (Instituto Nacional de Estatísticas, 2022).

In the Brazilian case, the reference year is 1870, when the "*Diretoria Geral de Estatísticas*" (General Statistics Directory) was founded. This organization was the first entity to centralize the Official Statistics

---

[1] BBC. Statistics reveal Britain's 'Mr and Mrs Average'. In: https://www.bbc.com/news/uk-11534042 . January, 29, 2019.

in the South American country. Two years late, in 1872, that organization ran the first Brazilian demographic Census (Diretoria Geral de Estatística 1874).

Considering the extensive history about Official Statistics, its operators, named in this study as National Statistics Offices (NSOs), adapted and evaluated their businesses across the times. Actually, the NSOs are dealing with a new round of changes into the business cycle, related to the adoption of Big Data technologies. Next section explains the NSOs business cycle.

Attributions and uses of the Official Statistics

According to Radermacher (2020), official statistics are essential to societies because they can be used to support public policies, business and investment decisions. Official statistics also offer to citizens the capability to verify if governments are deploying social welfare.

In this way, the National Statistics Offices plays a central role in the Official Statistics business. Radermacher (2020) highlights those offices as keeper of the uniformized production of Official Statistics, following effective and efficient patters. This perspective can be realized in the missions claims of each National Statistics Offices (NSOs). As examples, the organizational missions of National Statistics Offices from UK (ONS), Brazil (IBGE), and Portugal (INE) represents clear commitment with the society and its stakeholders:

"High quality data and analysis to inform the UK, improve lives and build the future." (ONS - Office for National Statistics)

"The Mission of Statistics Portugal is to produce, in an independent manner, high-quality official statistical information, relevant for the society, while promoting the coordination, the analysis, the innovation and the dissemination of the national statistical activity and ensuring integrated data storage." (INE - Instituto Nacional de Estatísticas)

"To portray Brazil by providing the information required to the understanding of its reality and the exercise of citizenship." (IBGE - Instituto Brasileiro de Geografia e Estatísticas)

After a simple verification with the use of tag cloud chart, it is possible to realize the bases of the Official Statistics are presented in those three NSOs (ONS, IBGE, INE). Figure 1 shows this tag cloud. The relevance of the words like "information", "quality" and "analysis" indicates a common basis for the NSOs define their missions. Those words can be translated to values or targets to be achieved by those organizations. In this way, it is necessary to explain that the NSOs follow a kind of common "constitutional bill", with the principles and guidelines to define the Official Statistics procedures.

Figure 1 – Relevant terms mentioned in the NSOs strategic missions' claims

In fact, this "constitutional bill" work as a list of limits and protocol into the NSOs can act. Everything out of this frame is considered misconduct by NSOs' pairs. This document is named as the "*Fundamental Principles of Official Statistics*" (United Nations, 2014).

Ten principles guide the operations of Official Statistic by NSOs. Therefore, those ten principles approach specific ten themes about de execution of the Official Statistics[2]. The first announces the relevance and the use of the Official Statistics. Next, the second principle approaches the trust preservation by the society on the Official Statistics, based on the scientific principles and professional ethics. Data interpretation and reinterpretation are the themes of the principles three and four. In the sequence, principle five approaches the sources of data, including surveys and administrative records. Individual privacy, regulations and laws are the themes of the principles six and seven. Regarding the cooperation in international level, principles eight, nine and ten not only propose but also encourages the partnerships between NSOs.

In order to guide and advance the use of Official Statistics, those principles work as starting point from NSOs develop their operations and targets. Indeed, the principles are not a self-center set of rules. United Nations (2022) and Radermacher (2020) remark the Official Statistics are being used to inform about the United Nations Milleniunm Development Goals (MDGs). In this case, surveys and administrative data support the measurement of the countries progress to achieve those goals.

So, it is possible to define that Official Statistics is a type of business (Radermacher 2020). As a type of business, it has the own production cycle. In this case, a data production cycle. There is also a business architecture to support the Official Statistics Business (Radermacher 2018). Both described by Radermacher (2020) and structed and aligned with the Fundamental Principles of the Official Statistics.

Figure 2 presents the data generation life cycle into Official Statistics. In this circuit is possible to realize the users of statistics, the society in general, as the spark of the demand for Official Statistics outcomes. Development, quality verification and communication/dissemination are stages in the Official Statistics life cycle.
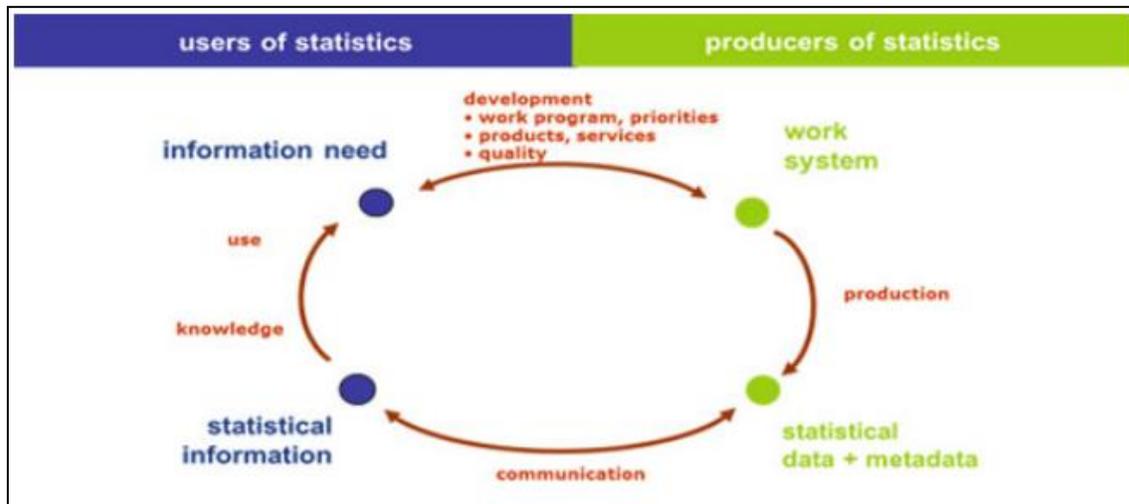

Figure 2 – Life cycle of the Official Statistics (Radermacher 2020)

It is relevant to remark the second green dot presents the raw material of Official Statistics, data and metadata. Just after the processing and the analysis of those elements, the statistical information can be disseminated.

---

[2] United Nations Statistic Divison. Fundamental Principles of Official Statistics. In: https://www https://unstats.un.org/unsd/dnss/hb/E-fundamental%20principles_A4-WEB.pdf. July, 1st, 2022.

The Official Statistical life cycle, associated with the Fundamental Principals, allows the development of the business architecture followed by the National Statistics Offices, and portraited by Radermacher (2018). Figure 3 presents this extensive scheme. It shows the value chain of the Official Statistics in the rectangle named "Production". Surveys and metadata are the core of this value chain. The methodology and the technology are highlighted in the domain of "Develop". Training, legal, organizational structure, ICT in all perspectives (capabilities, complexities, readiness, infrastructure) constitute the support line to run official statistics business. Communication is the final stage of this architecture. It is understood as a wide range of dissemination channels, since the IT capability to offer user access (Data warehouses and micro data access), until publishing (publications and press office).
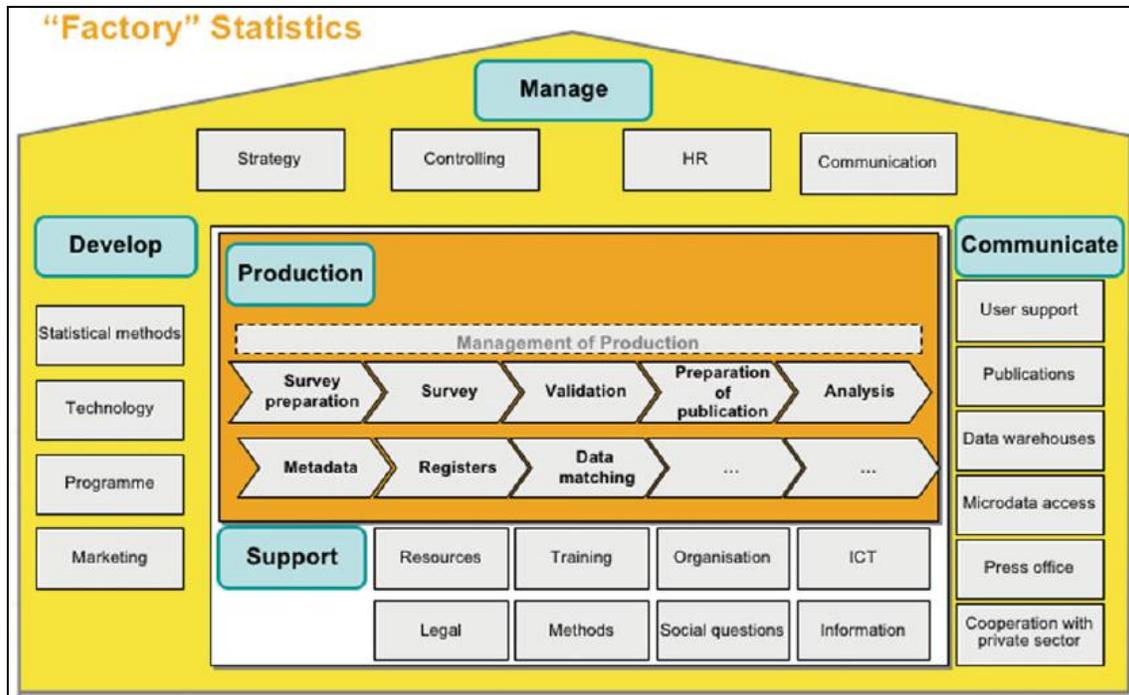

Figure 3 – Business Architecture of the Official Statistics (Radermacher 2018)

Moreover, all those attributions and uses related to Official Statistics are a descriptive contextualization. This context is where the Big Data adoption is in Official Statistics. Regarding the theme of this research is the adoption of Big Data by NSOs, it is mandatory to explain the core elements of Official Statistics and National Statistics Offices.

Not only survey and meta data integrates the data collection executed by NSOs. With a periodicity of a decade (Instituto Nacional de Estatísticas, 2022), the NSOs deploy a huge operation, the Demographic Census. This population national accounting take places simultaneously with the other surveys. The Census also become the base of future public policies (IBGE, 2022).

It is relevant to cite the NSOs' portfolios are similar, and excluding one or two specific national surveys, they present pattern types of research. The United Nations Statistic Division presents in its webpage (United Nations 2021) a set of data aggregated and compiled by United Nations Statistics Division (UNSD). This amount of data is sent to the UNSD by the NSOs. The patterns and the quality of the surveys make possible the data join and data association from different organizations.

In summary, the exchange of information and methodologies between the NSOs and USND with the NSOs characterizes a global organizational network, as remarked by Cardoso et al. (2021, p.3):

"As nodes of the NSOs network, these organizations have similarities regarding the organizational mission and operations. They respond to their countries' respective governments as well as the United Nations Statistics Division - UNSD (UNSD 2019). The joint operation of NSOs within UNSD forms the Global Statistical System (GSS). In other words, the GSS natural configuration is a global network organization (Mizruchi and Galaskiewicz 1994). On top of that worldwide structure, some NSOs and UNSD play central roles in operational standardization and technology innovation processes. In contrast, others GSS members act under a follower perspective. The GSS represents a Core-Periphery network standard (Kurt and Kurt 2020). Its members keep working as a network (The United Nations Statistical Commission 2020) related to Big Data Adoption. "

The NSOs and the GSS are facing a contemporary challenge/dilemma: innovation or obsolescence. Regarding the Official Statics are consumed by society, in the last two decades the citizens, companies, and other stakeholders demanded faster and faster data from GSS and NSOs. If those players do not deploy the data in real time, the stakeholders search related information in the competitor actors, such as the Tech Giants. Cardoso et al. (2021, p.3) highlighted about this scenario of threatens:

"As official data providers, NSOs direct their work processes, surveys, studies, and projections to guide and support public and private stakeholders' decision-making worldwide. NSOs can also be classified as data actors (Pappas et al. 2018). Nowadays, those organizations deal with the challenge of losing relevance to societies in the face of new data providers (Milan and Treré 2020). Those new players are Tech Giants (TGs) such as Google, Apple, Facebook, and Alibaba. Some of them are under governmental pressure after provoking distortions in prices of specific markets, such as tourism (The Independent 2021) and digital advertisement (Financial Times 2021)."

Then, after over four hundred years of evolution, Official Statistics and their operators, the NSOs, are under pressure for innovation. This is the context where the Big Data adoption is or should be deployed by the NSOs and GSS. However, this deployment is not well stablished. Considering the UNSD report during the COVID-19 pandemic (The United Nations Statistical Commission, 2020), less than the half of NSOs presented in that moment Big Data solutions in their core process. This is a threat discussed in the following subsection.

The threat of the Big Data adoption

Big Data adoption represents a threat to NSOs when this adoption in consider under a time context. Allin (2021) highlights the case of the Official Statistics in UK. The author indicates that the NSOs must redefine their role in the society in those Big Data times. So, the risk of obsolescence comes joined with the opportunity to renew the society's confidence and attention. In this way, Allin, (2021, p. 166) remarks

"National statistics offices are not just suppliers of statistics; they should also be providing answers to questions raised in society, recognizing that they have to compete for users' attention. By succeeding in doing so, they can confirm the value of official statistics to society."

This redefinition of the NSOs organizational role includes the innovation cycle. Many of the NSOs plan and work with a ten-years-innovation plan, based on the Census operation. Actually, the innovation demand is almost occurring in daily basis. In this way, the innovation cycle into NSOs must be reduced, maybe to yearly periods. It is not simple to promote this change. Considering the history of Official Statistics, this is a scientific field where the innovation took place in a gradual way, not in a disruptive one.

Additionally, the incapacity to work based on short-time projects is another relevant threat to NSOs. Despite of the NSOs work in long-term projects, maybe they are no longer well adapted to short-term projects. Bakici et al. (2022), concluded the Big Data adoption demands a project readiness from the adopter. This capability is also to operational and strategic projects. Following Bakici approach, Bilal et al. (2016) indicates the project readiness is the success key of the Big Data adoption in the construction industry.

There is a third risk related to the Big Data adoption by NSOs, the top management support. Baig et al. (2019, 2021) and Sun et al. (2018) concluded this is a critical risk factor in the adoption process. Both

authors indicate if the top managers sponsoring the proposal of innovation, the Big Data can be adopted. So, the NSOs will be well succeed if the top managers "buy" the idea and the project.

In summary, those three risks: (i) innovation cycle, (ii) short-term project readiness and (iii) top management support are just a sample of the risks related to NSOs and Big Data adoption. In fact, those three risks are very relevant, but isolated.

Conclusion

This Completed Research Paper discussed the trajectory of Official Statistics and analyzed the business architecture of the official data providers, the National Statistics Offices. According to the referenced papers and recent facts, NSOs are dealing with a dilemma: innovation or obsolescence. This dilemma can be translated into resistance or acceptance of Big Data technologies and methods.

In summary, the next few years are critical to maintaining GSS and NSOs as relevant organizations on the Global stage. If these organizations lose public relevance and are no longer considered by their stakeholders, the main asset of this business, trust, will be lost.

Future studies can be addressed to understand and list the risks inherent in adopting or avoiding of Big Data in the National Statistics Offices' business cycle. Qualitative or quantitative research will aggregate knowledge to this study about those relevant organizations.

References

Allin, P. (2021). "Opportunities and challenges for official statistics in a digital society," *Contemporary Social Science*, *16*(2), 156–169. https://doi.org/10.1080/21582041.2019.1687931

Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2019). "Big data adoption: State of the art and research challenges," *Information Processing and Management*, *56*(6). https://doi.org/10.1016/j.ipm.2019.102095

Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2021). "A Model for Decision-Makers' Adoption of Big Data in the Education Sector," *Sustainability*, *13*(24), 13995. https://doi.org/10.3390/su132413995

Bakici, T., Nemeh, A., & Hazir, O. (2022). "Big Data Adoption in Project Management: Insights From French Organizations," *IEEE Transactions on Engineering Management*, *PP*, 1–15. https://doi.org/10.1109/TEM.2021.3091661

Bardin, L. (2016). *Análise de Conteúdo,* 1st Ed, Almedina Brasil.

BBC. (2019). *Statistics reveal Britain's "Mr and Mrs Average."* https://www.bbc.com/news/uk-11534042

Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., Owolabi, H. A., Alaka, H. A., & Pasha, M. (2016). "Big Data in the construction industry: A review of present status, opportunities, and future trends," *Advanced Engineering Informatics*, *30*(3), 500–521. https://doi.org/10.1016/j.aei.2016.07.001

Cardoso, F. dos S., Varajão, J., & Carvalho, A. (2021). "The adoption of Big Data technologies - A challenge for National Statistics Offices," *AMCIS 2021 Proceedings*. https://aisel.aisnet.org/amcis2021/data_science_decision_support/data_science_decision_support/13

Cresswell, J., & Cresswell, J. D. (2018). *Research Design - Qualitative, Quantitative, and Mix Methods Approaches*, 5th Ed, SAGE Publications, Inc.

Diretoria Geral de Estatística. (1874). *Recenseamento do Brazil em 1872,* Editor: Leuzinge, G. https://biblioteca.ibge.gov.br/visualizacao/livros/liv25477_v1_br.pdf

IBGE. (2022). *Comitê de Estatísticas Oficias*. https://ces.ibge.gov.br/apresentacao/portarias/200-comite-de-estatisticas-sociais/base-de-dados/1146-censo-demografico.html

Instituto Nacional de Estatísticas. (2022). "História dos Censos em Portugal," *Censos Em Portugal*. https://censos.ine.pt/xportal/xmain?xpid=INE&xpgid=censos_historia_portugal

Kotz, S. (2005). "Reflections on Early History of Official Statistics and a Modest Proposal for Global Coordination," *Journal of Official Statistics*, *21*(2), 139–144.

Quetelet, A. (2013). *A Treatise on Man and the development of this faculties, Editor:* T. Smibert, Cambridge University Press. http://www.cambridge.org/9781108064422

Radermacher, W. (2018). *"ESS vision and ways for cooperation," Workshop on strategic developments in business* (Issue January). https://doi.org/10.13140/RG.2.2.32812.77448

Radermacher, W. J. (2020). *Official Statistics 4.0*. Springer International Publishing. https://doi.org/10.1007/978-3-030-31492-7

Royal Statistical Society. (2022). *History*. About. https://rss.org.uk/about/history/

Stigler, S. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Harvard University Press.

Sun, S., Cegielski, C. G., Jia, L., & Hall, D. J. (2018). "Understanding the Factors Affecting the Organizational Adoption of Big Data", *Journal of Computer Information Systems*, *58*(3), 193–203. https://doi.org/10.1080/08874417.2016.1222891

The United Nations Statistical Commission. (2020). *Background document for the Note by the Secretary-General transmitting the report of the Global Working Group on Big Data for Official Statistics*, United Nation Statistics Division, United Nations Statistics Division.

UK Statistics Authority. (2022). *Types of Official Statistics*. About the Authority. https://uksa.statisticsauthority.gov.uk/about-the-authority/uk-statistical-system/types-of-official-statistics/

United Nations. (2014). "Fundamental Principles of Official Statistics", In *Sixty-eighth session of General Assembly* (Vol. 9, Issue January).

United Nations. (2021). *UN Data*. http://data.un.org/Explorer.aspx

United Nations. (2022). *Millennium Developtment Goals and Indicators*. https://mdgs.un.org/unsd/mdg/default.aspx

Van Campenhoudt, L., Marquet, J., & Quivy, R. (2017). *Manuel de recherche en sciences sociales,* 5th Edition, Dunod.